

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 11, 2003, 06:24:06 ; Search time 4227 Seconds
(without alignments)
4461.637 Million cell updates/sec

Title: US-09-043-944-1
Perfect score: 2378
Sequence: 1 MPSTRQEGGGADAEHTV.....YYINSLFLPFLCIINFSIIS 461

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 288711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+p2n.model -DEV=xl
-Q=/cgn2.1/USPTO spoil p/US09043944/runat_11122003_062400_3671/app query.fasta_1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=rg -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcg -NOR=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09043944 @CGN_1.1.4958 @runat_11122003_062400_3671 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*

RESULT 1

- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2362	99.3	1461	3	CEU35660
2	2263	95.2	1335	3	AF171064
3	2052.5	86.3	27102	3	U41540
4	1009.5	42.5	1836	5	AY043492
5	997	41.9	1896	4	AF038936
6	993.5	41.8	1488	6	AF086330
7	993.5	41.8	2764	6	AR095621
8	993.5	41.8	2765	6	AX244146
9	993.5	41.8	2765	6	AX244648
10	993.5	41.8	2765	9	HUMS182R
11	993.5	41.8	2765	11	G27112
12	993.5	41.8	2791	6	AR060154
13	993.5	41.8	2791	6	AR087309
14	993.5	41.8	2791	6	AR134496
15	993.5	41.8	2791	6	AR256768
16	993.5	41.8	2792	6	AR144597
17	993.5	41.8	2811	9	BC011729
18	992	41.7	1764	3	DMU78084
19	989.5	41.6	2791	6	AR060024
20	989.5	41.6	2791	6	AR087179
21	989.5	41.6	2791	6	AR134366
22	989.5	41.6	2791	6	AR144467
23	989.5	41.6	2791	6	AR256638
24	989.5	41.6	3056	9	HUMPS1MRNA
25	989.5	41.6	3087	6	AR060026
26	989.5	41.6	3087	6	AR087181
27	989.5	41.6	3087	6	AR134368
28	989.5	41.6	3087	6	AR144469
29	989.5	41.6	3087	6	AR256640
30	989	41.6	1962	6	AR060155
31	989	41.6	1962	6	AR087310
32	989	41.6	1962	6	AR134497
33	989	41.6	1962	6	AR256769
34	989	41.6	1964	6	AR144598
35	989	41.6	1964	10	MUS51PR
36	989	41.6	2016	10	BC030409
37	989	41.6	2681	6	AG3557
38	989	41.6	2681	6	AR271266
39	988.5	41.6	1809	3	DMU77934
40	988.5	41.6	1894	3	AF017024
41	988.5	41.6	1895	6	AR060183
42	988.5	41.6	1895	6	AR087338
43	988.5	41.6	1895	6	AR134525
44	988.5	41.6	1895	6	AR256797
45	988.5	41.6	1914	6	AR082573

ALIGNMENTS

CEU35660	1461 bp	mRNA	linear	INV 23-JAN-1996
LOCUS				
DEFINITION	Caenorhabditis elegans membrane protein (sel-12) mRNA, complete cds.			
ACCESSION	U35660			
VERSION	U35660.1	GI:1109793		
KEYWORDS	Caenorhabditis elegans			
SOURCE	Caenorhabditis elegans			
ORGANISM	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis. (bases 1 to 1461)			
REFERENCE	Levitani, D. and Greenwald, I.			
AUTHORS	Facilitation of lin-12-mediated signalling by sel-12, a			
TITLE	Caenorhabditis elegans S182 Alzheimer's disease gene			
JOURNAL	Nature 377 (6547), 351-354 (1995)			
MEDLINE	96032531			
PUBMED	7566091			
REFERENCE	2 (bases 1 to 1461)			
AUTHORS	Levitani, D.			
TITLE	Direct Submission			
JOURNAL	Submitted (07-SEP-1995) Diane Levitan, Biochemistry, Columbia			
COMMENT	University, 630 W. 168th St., New York, NY 10032, USA			
FEATURES	On Dec 9, 1995 this sequence version replaced gi:1000566.			
source	Location/Qualifiers			
	1..1461			
	/organism="Caenorhabditis elegans"			
	/mol_type="mRNA"			
	/strain="Bristol var N2"			
	/db_xref="taxon:6239"			
	/map="IGX"			
gene	1..1461			
	/genes="sel-12"			
CDS	1..1386			
	/notes="multiple membrane spanning domains"			
	/codon_start=1			
	/product="membrane protein"			
	/protein_id="AAA85511.1"			
	/db_xref="GI:1109794"			
	/translation="MPSRRQEGGGADAETHVYTNLTITNRNSQEDENVVEAEFLK			
	YGASHVHLFPVSLCMALVFTMTNTITFYSONGRHLLYTPPRETDSIVKGLMSL			
	GNALMVCVVMATVLLVYKYKFKLHGLWLVSSFLLLFTIYVQVLEKSFV			
	SPSALLVPLGLNYGVLMCHWKGRLRQPFYLLTMSALMALVFKYLPWVWFV			
	LFVSSWDVLVLTGPKURLYVETAQERNEPIFPALIVSSGVIPYVLVTAVENTTD			
	PREPSSDNTAFPEASCSETPKPKVKRIPOKVOIESNTTASTTQNSGVVER			
	ELAAERTVQDANFHEHEERGVKLGDFIFYSVLGKASSYFDWNTTIACYVAIL			
	IGLCFTLVLLAVFKRALPALQFPSPDFTTFVPAGSSPHLLHKSLSKSYVINSFLP			
	FLCIINFSIIS"			
polyA_site	1461			
	/genes="sel-12"			
BASE COUNT	379 a	309 c	310 g	463 t
ORIGIN	/note="20 A residues"			
Alignment Scores:				
Pred. No.:	2,19e-199	Length:	1461	
Score:	2362.00	Matches:	459	
Percent Similarity:	99.57%	Conservative:	0	
Best Local Similarity:	99.57%	Mismatches:	2	
Query Match:	99.33%	Indels:	0	
DB:	3	Gaps:	0	
US-09-043-944-1 (1-1461) x CEU35660 (1-1461)				
QY	1	MetProSerThrArgArgGlnGlnGluGlyGlyValAlaAspAlaGluThrHisThrVal	20	
Db	1	ATGCGCTCCACAGAGGACACACGAGGCGGAGGTCAGATCGCGAAACATACCGTT	60	
QY	21	TyrGlyThrAsnLeuLeuThrAsnArgAsnSerGlnGluAspGluAsnValValGluGlu	40	
Db	61	TACGGGTACAAATCTGATAACAAATCGAATAGCAGAGAGAGAGAGAGAGAGAGAG	120	
QY	401	LeuLeuAlaValPheLysArgAlaLeuProAlaLeuGlnPheProPheSerProAspSer	420	
QY	41	AlaGluLeuLysTyrGlyAlaSerHisValIleHisLeuPheValProValSerLeuCys	60	
Db	121	GCGGAGCTGAAATACGAGGACATCTCAGTTATTCACTATTGTGCGCGGTGCTACTATGC	180	
QY	61	MetAlaLeuValValPheThrMetAsnThrIleThrPheTyrSerGlnAsnAsnGlyArg	80	
Db	181	ATGGCTCTGGTTCTGTTTATACGATGAACACAGATTACGTTTATAGTCAAAACAATGGAAGG	240	
QY	81	HisLeuLeuSerHisProPheValArgGluThrAspSerIleValGluLysGlyLeuMet	100	
Db	241	CAITTTACTATACACTCTCTTTTCCGGGAAACAGACAGTATCGTTGAGAGGGAATTCATG	300	
QY	101	SerLeuGlyAsnAlaLeuValMetLeuCysValValValLeuMetThrValLeuLeuIle	120	
Db	301	TCACTTTGGAAATGCTCTCGTCATGTTGTGGTGGTCTGTCGATGACAGTTCTGCTGATT	360	
QY	121	ValPheTyrLysTyrLysPheTyrLysLeuIleHisGlyTrpLeuIleValSerSerPhe	140	
Db	361	GTITTTCTATAAATACAAAGTTTATAAGCTTATTTCATGGATGGCTTATTGTGACAGATTTT	420	
QY	141	LeuLeuLeuPheLeuPheThrThrIleTyrValGlnGluValLeuLysSerPheAspVal	160	
Db	421	CTTCTCTCTTTCTCTATCTACTACATCTATGTCAGAAAGTTCTGAAAAGTTTCGATGTG	480	
QY	161	SerProSerAlaLeuLeuValLeuPheGlyLeuGlyAsnTyrGlyValLeuGlyMetMet	180	
Db	481	TCCTCCAGCGCACTATTGTTTGTGTTGTTGACATGAGTATCGAGTTCTCGGAATCATG	540	
QY	181	CysIleHisTrpLysGlyProLeuArgLeuGlnGlnPheTyrLeuIleThrMetSerAla	200	
Db	541	TGTATACATTGAAAGGTCATTTGCGTCTGCAACAGTTCTACCTTATTACAAATGTCTGCA	600	
QY	201	LeuMetAlaLeuValPheIleLysTyrLeuProGluTyrThrValTrpPheValLeuPhe	220	
Db	601	CTATGCGCTCTGCTGCTTTTATCAAGTACCTACCAAGATGGACTGTGGTGTGTGCTGTTT	660	
QY	221	ValIleSerValTrpAspLeuValAlaValLeuThrProLysGlyProLeuArgTyrLeu	240	
Db	661	GTATCTCGGTTTGGATCTGGTTGCGGTGCTCACACCAAGAGACCATTCAGATATTG	720	
QY	241	ValGluThrAlaGlnGluArgAsnGluProIlePheProAlaLeuIleTyrSerSerGly	260	
Db	721	GTGGAAATCGCACAGAGAGAAACAGCCAAATTTCCCGCGCTGATTATTTCGTCTGGA	780	
QY	261	ValIleTyrProTyrValLeuValThrAlaValGluAsnThrThrAspProArgGluPro	280	
Db	781	GTCATCTATCCCTACGTTCTTGTCTGAGTTGAAACACGACACCCCGGTGAACCG	840	
QY	281	ThrSerSerAspSerAsnThrSerThrAlaPheProGlyGluAlaSerCysSerSerGlu	300	
Db	841	ACGTGTCAGACTCAAAATACTTCTACAGCTTTTCTGGAGAGCGGAGTTGTTTCATCTGAA	900	
QY	301	ThrProLysArgProLysValLysArgIleProGlnLysValGlnIleGluSerAsnThr	320	
Db	901	ACGCCAAAACGGCCAAAAGTGAACGAATTCCTCAAAAAGTGCAAAATCGAATCGAATCT	960	
QY	321	ThrAlaSerThrThrGlnAsnSerGlyValArgValGluArgGluLeuAlaGluArg	340	
Db	961	ACAGCTTCACAGCACACAAAACCTCGAGTAAAGGTGGAGCGGAGCTACTCTCTGAGAGA	1020	
QY	341	ProThrValGlnAspAlaAsnPheHisArgHisGluGluGluArgGlyValLysLeu	360	
Db	1021	CCAACTGTACAAGACGCCAAATTTTCACAGGCACGAGAGAGAGAGAGAGAGAGAGAG	1080	
QY	361	GlyLeuGlyAspPheIlePheTyrSerValLeuLeuGlyLysAlaSerSerTyrPheAsp	380	
Db	1081	GGTCTGGGCGACTTCATTTTCTACTCTGTTCTCTCTCGGACAGGCTTCATCGTACTTTGAC	1140	
QY	381	TrpAsnThrThrIleAlaCysTyrValAlaIleLeuIleGlyLeuCysPheThrLeuVal	400	
Db	1141	TGGAACACAGCATATCGCTGTTGTTATGGCCATCTTATCGGCTCTCTGCTTACTCTTGTG	1200	

```

|||||
1201 CTGGCTCGCGCTTTCAACAGCAGACTCCCGGCTCTGCAATTTCCATTTCTCGGACTCA 1260
Qy 421 PhePheThrPheValProAlaGlySerSerProHisLeuLeuHisLeuLeuHisLeuLeuVal 440
Db 1261 TTTTCTACTTTTGTACCGCTGGATCATCACCCCATTTCTTACACAAGTCTCTCAAAAGT 1320
Qy 441 ValTyrTyrIleAsnSerLeuPheLeuProPheLeuCysIleIleAsnPheSerIleIle 460
Db 1321 GTTTATTATTAATTCTCTGTTTGTGCAATTTCTTGTGCATCATCAACTTTTCGATTATA 1380
Qy 461 Ser 461
Db 1381 TCT 1383

RESULT 2
AF171064 1335 bp mRNA linear INV 25-OCT-2002
LOCUS Caenorhabditis elegans presenilin SEL-12 (sel-12) mRNA, complete
DEFINITION
ACCESSION AF171064 GI:5759132
VERSION AF171064.1
KEYWORDS Caenorhabditis elegans
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;
Rhabditoidea; Rhabditiidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 1335)
AUTHORS Wittenburg,N., Eimer,S., Lakowski,B., Rohrig,S., Rudolph,C. and
Baumeister,R.
TITLE Presenilin is required for proper morphology and function of
neurons in C. elegans
JOURNAL Nature 406 (6793), 306-309 (2000)
MEDLINE 20372200
PUBMED 10917532
REFERENCE 2 (bases 1 to 1335)
AUTHORS Eimer,S. and Baumeister,R.
TITLE Direct Submission
JOURNAL Submitted (22-JUN-1999) Genzentrum, LMU Munich, Feodor-Lynen-Str.
25, Munich 81377, Germany
FEATURES
source
1..1335
Location/Qualifiers
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/chromosome="X"
1..1335
/gene="sel-12"
1..1335
/gene="sel-12"
/function="facilitator of Notch receptors signaling"
/note="membrane protein; similar to Homo sapiens PS1 and
PS2; similar to SEL12 deposited in GenBank Accession
Number AAA85511"
/codon_start=1
/product="presenilin SEL-12"
/protein_id="AAD50991.1"
/db_xref="GI:5759133"
/translation="MPSTRQEGGDAETHVTGNTLITNRNSQEDENVVBAELK
YGASHVILFVPSLCAALVFTWNTITFYSONNGRHLLVTPFVRETDISIVEKGLMSL
GNALVLCVVVLMTLLVLFVYKPKYKLIHGKLVSSFLLLFLFTTIYQVBLKSFV
SPSALLVFLGNGVGLGMCIHKKGRLQQLQFIIITMSALMLVFIKILPEWTFV
LFVLSMDLVAFLTPKGRXLIVETAQERNEPIFPALYSSGVIPYVLVAVENTTD
PREPTSDSNTSTAPGSCSSSTPKPKVRIPOKQVIESNTASTTQNSGVYRVER
ELAAERPTVQDANFHRHEERGVKGLGDFIYSLVILGKASSYFDWNTTIACYVAIL
IGLCFTLLVLAFFKRALPALPISIFSGLIFFYFCTRWIITPFTVQVSQKLLY"
BASE COUNT 345 a 291 c 301 g 398 t
ORIGIN
Alignment Scores: 1.11e-190 Length: 1335
Pred. No.: 2263.00 Matches: 442
Score:

```

```

Percent Similarity: 99.33% Conservative: 0
Best Local Similarity: 99.33% Mismatches: 2
Query Match: 95.16% Indels: 1
DB: 3 Gaps: 0
US-09-043-944-1 (1-461) x AF171064 (1-1335)

Qy 1 MetProSerThrArgArgGlnGlnGluGlyGlyGlyValaAspAlaGluThrHisThrVal 20
Db 1 ATGCCTTCACAAAGAGAGACACAGAGGGCGGAGGTGCAGATGCGGAAACACATACCGTT 60
Qy 21 TyrGlyThrAsnLeuIleThrAsnArgAsnSerGlnGluAspGluAsnValValGluGlu 40
Db 61 TACGGGTACAAATCTGATAACAAATCGAATGCCAAGAAGACGAAATGTGTGGAAGAA 120
Qy 41 AlaGluLeuLysTyrGlyAlaSerHisValIleHisLeuPheValProValSerLeuCys 60
Db 121 GCGGAGCTGAATAACGAGCATCTACGTATTATTCATATTGTGCGCGGTGTCACATATGC 180
Qy 61 MetAlaLeuValPheThrMetAsnThrIleThrPheTyrSerGlnAsnAsnGlyArg 80
Db 181 ATGGCTCTGGTTGTTTACGATGAACAGATTACGTTTTATATAGTCAAAACAATGAAGG 240
Qy 81 HisLeuLeuSerHisProPheValArgGluThrAspSerIleValGluLysGlyLeuMet 100
Db 241 CATTTACTATATACCTCTTTTTCGCGGAAACAGACAGTATCGTTTGAGAAGGATTTGATG 300
Qy 101 SerLeuGlyAsnAlaLeuValMetLeuCysValValValLeuMetThrValLeuLeuIle 120
Db 301 TCACCTTGGAAATGCTCTCGTCATGTTGTGCGTGGTCTGTCATGACAGTTTCTGCTGATT 360
Qy 121 ValPheTyrLysTyrLysPheTyrLysLeuIleHisGlyTyrLeuIleValSerSerPhe 140
Db 361 GTTTTCTATAAATACAAGTTTATAGCTTATTATGATGATGCTTATTGTCAGCAGTTT 420
Qy 141 LeuLeuLeuPheLeuPheThrThrIleTyrValGlnGluValLeuLysSerPheAspVal 160
Db 421 CTTCTCTCTTTCTTCTATTCTACATAATCTATGTGCAAGAAGTTCTGAAAAGTTTCGATGTG 480
Qy 161 SerProSerAlaLeuLeuValLeuPheGlyLeuGlyAsnTyrGlyValLeuGlyMetMet 180
Db 481 TCTCCACAGGCACTATTGGTTTGTGACTGGGTAACTATGGAGTTCTCGGAATGATG 540
Qy 181 CysIleHisTyrLysGlyProLeuArgLeuGlnGlnPheTyrLeuIleThrMetSerAla 200
Db 541 TGTATACATTGCAAGGTCCATTGGCTCTGCAACAGTTCTACCTTATTACAATGCTGCA 600
Qy 201 LeuMetAlaLeuValPheIleLysTyrLeuProGluTyrThrValTyrPheValLeuPhe 220
Db 601 CTAATGGCTCTGGTCTTTTCAAGTACCTACAGAAATGGACTGTGTGGTTTGTGCTGTT 660
Qy 221 ValIleSerValTyrAspLeuValAlaValLeuThrProLysGlyProLeuArgTyrLeu 240
Db 661 GTTATCTCGTTCGGATCTGGTTCGGTGTCTCACCAAAAGGACCATTTGAGATATTG 720
Qy 241 ValGluThrAlaGlnGluArgAsnGluProIlePheProAlaLeuIleTyrSerSerGly 260
Db 721 GTGAAACATGACAGAGAGAAAACGAGCAATTTTCCCGGCGCTGATTATTTCGTCGA 780
Qy 261 ValIleTyrProTyrValLeuValThrAlaValGluAsnThrThrAspProArgGluPro 280
Db 781 GTCATCTATCCCTACGTTCTTTGTACTGCGGTGAAAAACACAGACACACCCCGTCAACCG 840
Qy 281 ThrSerSerAspSerAsnThrSerThrAlaPheProGlyGluAlaSerCysSerSerGlu 300
Db 841 ACGTCGTCAGACTCAATACTTCTACAGCGTTTCTGAGAGGCGAGTTGTTCATCTCAA 900
Qy 301 ThrProLysArgProLysValLysArgIleProGlnLysValGlnIleGluSerAsnThr 320
Db 901 ACGCAAAACGGCCAAAGTGAAACGAATTCCTCAAAAGTCAATTCGAATCGAATACT 960
Qy 321 ThrAlaSerThrThrGlnAsnSerGlyValArgValGluArgGluLeuAlaGluArg 340
Db |||||

```

Db 961 ACAGTCTCAACGACACAAAACCTCTGGAGTAAAGGTCGGAACGGAGCTAGCTGCTGAGAGA 1020

Qy 341 ProThrValGlnAspAlaasnPheHisArgHisGluGluGluGluArgGlyVallysLeu 360
|||||

Db 1021 CCAACTGTACAGAGCGCCAAATTTTCACAGGCACGAAAGAGAGAGAGGTGTGAAACTT 1080

Qy 361 GlyLeuGlyAspPheIlePheTyrSerValLeuLeuGlyLysAlaSerSerTyrPheAsp 380
|||||

Db 1081 GGTCTGGGCGACTCATTTTCTACTCTGTCTCCCTCGGCAAGGCTTCATCGTACTTTGAC 1140

Qy 381 TtpAsnThrThrIleAlaCysTyrValAlaIleLeuIleGlyLeuCysPheThrLeuVal 400
|||||

Db 1141 TGGACACGACTATCGCCTGTTATGGGCATCTTATCGTCTCGTCTTCACTCTTGTTC 1200

Qy 401 LeuLeuAlaValPheLysArgAlaLeuProAlaLeu-GlnPheProPheSerProAspSe 420

Db 1201 CTGCTCGCGCTCTTCAAGAGCACTCCCGGCTCTGCCAATTTCCATTTCCTCGGACTC 1260

Qy 420 rPhePheThrPheValProAlaGlySerSerProHisLeuLeuHisLysSerLeuLysSe 440
|||||

Db 1261 ATTTTCTACTTTGTACCCGCTGGATCATCAACCCATTTGTTACACAGTCTCTCAAAAG 1320

Qy 440 rValTyrTyrIle 444
|||||

Db 1321 TGTATTATATAT 1333

RESULT 3

LOCUS U41540 27102 bp DNA linear INV 31-MAY-2003

DEFINITION Caenorhabditis elegans cosmid F35H12, complete sequence.

ACCESSION U41540

VERSION U41540.1 GI:1109862

KEYWORDS HTG.

SOURCE Caenorhabditis elegans

ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
1. (bases 1 to 27102)
Wilson, R.

REFERENCE
AUTHORS
TITLE
Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)

JOURNAL MEDLINE
PUBMED 99059613

REFERENCE
AUTHORS
TITLE
The sequence of C. elegans cosmid F35H12
Unpublished (2001)

JOURNAL
REFERENCE
AUTHORS
TITLE
Waterston, R.
Direct Submission

JOURNAL
REFERENCE
AUTHORS
TITLE
Submitted (24-APR-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 27102)
Waterston, R.
Direct Submission

JOURNAL
REFERENCE
AUTHORS
TITLE
Submitted (16-NOV-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
5 (bases 1 to 27102)
Waterston, R.
Direct Submission

JOURNAL
REFERENCE
AUTHORS
TITLE
Submitted (22-MAY-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
6 (bases 1 to 27102)
Waterston, R.
Direct Submission

JOURNAL
REFERENCE
AUTHORS
TITLE
Submitted (18-JUN-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
7 (bases 1 to 27102)
Waterston, R.

TITLE
JOURNALREFERENCE
AUTHORS
JOURNALREFERENCE
AUTHORS
JOURNAL

COMMENT

FEATURES
source

gene

CDS

Direct Submission
Submitted (19-NOV-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
8 (bases 1 to 27102)
Waterston, R.
Direct Submission
Submitted (19-APR-2003) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
9 (bases 1 to 27102)
Wilson, R.
Direct Submission
Submitted (31-MAY-2003) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: submissions@watson.wustl.edu and jesus@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this cosmid sequence and its analysis see:
<http://www.wormbase.org/db/seq/sequence?name=F35H12;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is ZC13, 200 bp overlap; the 3' cosmid is F53H8, 200 bp overlap. Actual start of this cosmid is at base position 1 of F35H12; actual end is at 12696 of F53H8.

NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yuji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFome cloning project (<http://worldb.dfci.harvard.edu/>), similarity to other proteins from BlastX analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

Location/Qualifiers
1..27102
/organism="Caenorhabditis elegans"
/mol_type="Genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="X"
/clone="F35H12"
2463..3251
/gene="F35H12.6"
join(2463..2552,3063..3251)
/gene="F35H12.6"


```

/standard name="F35H12.6"
/notes="contains similarity to Homo sapiens Glycophorin E precursor Glycophorin E precursor; ENSEMBL:ENSP00000305872; coded for by the following C. elegans cDNAs: Yk1092f08.3"
/codon_start=1
/product="Hypothetical protein F35H12.6"
/protein_id="AAM54184.1"
/db_xref="GI:21450554"
/db_xref="WormBase:F35H12.6"
/translation="MESYSGQVSESDSYSHYSNLTVQNLGAGVPMWAIWSFVICAV LIAVLALLDYCVRIRNALGNTCTKARQKRGTINAAHQKRSSTNMLNV"
5010. .7295
/gene="sel-12"
join(5010..5063,5134..5324,5629..5838,5884..6203,6265..6348,6398..6604,7027..7295)
/gene="sel-12"
/standard name="F35H12.3"
/notes="contains similarity to Pfam domain PF01080 (Presenilin); coded for by the following C. elegans cDNAs: Yk216el.3, Yk21d3.3, Yk21d3.5, Yk231a7.3, Yk231a7.5, Yk400e8.3, Yk400e8.5, Yk452b9.3, Yk452b9.5, Yk499e3.3, Yk573h4.3, Yk600e12.3, Yk674e3.3, AF171064, U35660, OSTF019E11.1, OSTR019E11.1"
/codon_start=1
/product="Suppressor/enhancer of lin-12 protein 12"
/protein_id="AAK39230.1"
/db_xref="GI:13775416"
/db_xref="WormBase:F35H12.3"
/translation="MPSTRRQEGGADAEHTVYGTNLITNRNSQEDENVVEAEELK YGASHVHLFVPSVCLMALVFTMTVITFYSONGRHLLYTPFRETDSIVKGLMSL GNALVMLCVLMTVLTVYKYKFKLHGWLIVSSFLLEFLFTIYQVQLKSFV SPALLVFLGNYGVGLNMCVHKGPLRLQOPLYLITGSAALMALPIKIPLEWTVFV LFVSVWDLVAVLTPGKPLVYETAQRNEPFIPIALYSSGVYIPVYLTVAVENTD PREPTSSDNTSTAFPGESCSSETPKPKVKRIIPQKQIESNTASTTQNSGVVER ELAERPVTQDANFHRHBEERGVKLGDFIFYSVLGKASSYFDMWNTTIACYVAIL IGLCTLLVLAVKRALPALPISIFSLGIFYFCTRWIITPFTVQVSKCLLY"
7533. .9207
/gene="F35H12.5"
join(7533..7580,7730..7867,7922..8227,8270..8349,8659..8744,8792..9207)
/gene="F35H12.5"
/standard name="F35H12.5a"
/notes="contains similarity to Interpro domains IPR000734 (Lipase), IPR000379 (Esterase/lipase/thioesterase, active site); coded for by the following C. elegans cDNAs: Yk703h10.5"
/codon_start=1
/product="Hypothetical protein F35H12.5a"
/protein_id="AAK39231.2"
/db_xref="GI:15950403"
/db_xref="WormBase:F35H12.5a"
/translation="MFSRRHILSNVRPIIKNLDLTRGFSDRKRYFDLNDLIKTVNVT YKTGNLEETTVESKYVDVTPSSNTKGTVTLSGPGTHNDFKYMKSFFQKKIRLI CTNYPGSEFVGGHNSYTNQDRNSYMKSLMETLELKNVRLIIMGHSRGGENALQLT SMLNDENWPLGVAMWINSPGFAPKHGISKRMGTINFILSLIKRHNKINSILPILH YFYNLIGLVSHGKVAALILPMQTFAPDEQKLSIDDIRAKPGIRAFYGYGSKDFLI DEHQSEVAMTFSESDHVTISNKQDAEKAIKARKSFTTGQFVTANFKEGHFLQKT YPEFIVEVDSIFDADKEVDTKI"
Join(8003..8220,8269..8349,8659..8744,8792..9207)
/gene="F35H12.5"
/standard name="F35H12.5b"
/notes="contains similarity to Interpro domains IPR000734 (Lipase), IPR000379 (Esterase/lipase/thioesterase, active site); coded for by the following C. elegans cDNAs: OSTF212H6.1, OSTR212H6.1"
/codon_start=1
/product="Hypothetical protein F35H12.5b"
/protein_id="AAP46284.1"
/db_xref="GI:31249887"
/db_xref="WormBase:F35H12.5b"
/translation="MKSPFEQKIRLICTNYPGSEFVTVGLHNSYTNQDRNSYMKSLM ETLELKNVRLIIMGHSRGGENALQLTSMNDENWPLGVAMWINSPGFAPKHGISKRMG

```

```

TINFIISLIKRNKNTINSILPHILHYFNLLIGLRVSHGKVAALILPMQTFAPDEQK
LSIDDLRAKPGIRAFYGYGSKDFLIDHSESEVAMTFSEHDHYVINSKQDAEAIKKA
RKSFTTGKQFVTANFKEGHFLQKTYPEFIVEVDSIFDADKEVDTKI"
complement(9443..13031)
/gene="F35H12.4"
complement(join(9443..9627,10392..10517,10573..10772,10828..11054,11147..11335,11383..11511,11615..11845,11895..12010,12526..12728,12776..12893,12941..13031))
/gene="F35H12.4"

Alignment Scores:
Pred. No.: 3 33e-170 Length: 27102
Score: 2052.50 Matches: 459
Percent Similarity: 58.82% Conservative: 1
Best Local Similarity: 58.70% Mismatches: 1
Query Match: 86.31% Indels: 322
DB: 3 Gaps: 7

US-09-043-944-1 (1-461) x U41540 (1-27102)
QY 1 MetProSerThrArgArgGlnGlnGlnGlyGlyGlyAlaAspAlaGluThrHis----- 18
Db 5010 ATGCCTTCCACAGGAGACACACAGGAGGGGGAGGTGCAGATCGCGAAGACACATGAAGT 5069
QY 18 ----- 18
Db 5070 TATTTAGACATTTATTTTCTCAAGAACTAAATGTTAAAAATGCTACAATGCATTTGTT 5129
QY 19 ----ThrValTyrGlyThrAsnLeuIleThrAsnArgAsnSerGlnGluAspGluAsnVa 37
Db 5130 TCAGACCGTTTACGGTACAAATCTGATACAATCGGAATAGCAAGAGACGAAAATGT 5189
QY 37 lValGluGluAlaGluLeuLysTyrGlyAlaSerHisValIleHisLeuPheValProVa 57
Db 5190 TGTGGAAGAGCGGAGCTGAATACGAGCATCTCAGTATTTCATCTATTGTGCGCGT 5249
QY 57 lSerLeuCysMetAlaLeuValValPheThrMetAsnThrIleThrPheTyrSerGlnAs 77
Db 5250 GTCACTATGCATGGCTCTGGTTGTTTACGATGAACACAGATACGTTTATAGTCAAA 5309
QY 77 PAsnGlyArgHisLeu----- 82
Db 5310 CAATGGAAGCAATTTGTAAGTTTCTAAGAANTGATGATTAATAATATTGATTTGTT 5369
QY 83 -----LeuSer----- 84
Db 5370 TTATCAATTTGCATCTGTGCATCGCACTCTTTGTGTCAGTGAATAATTTTGGTCAGT 5429
QY 84 ----- 84
Db 5430 GCAAAATAATTTTGTGTCAGTCATTTGTTATTTGTCAGTCAGTCAGTCAGTCAGTC 5489
QY 84 ----- 84
Db 5490 TTAACATTTTGTGCGTTTAAATTTTACTCAATTTTCTATCAATTTTCTTTGGAAAAA 5549
QY 84 ----- 84
Db 5550 GTTGAAGATTTACTCTGGAATTTTGAATAAATGTAATGGAATAATCAATCAACACA 5609
QY 85 -----His-ProPheValArgGluThrAspSerIleValG 96
Db 5610 AACTTTGAAATTTATTTTCAGACTATACACTCTTTTGTCCGGGAAACACAGATGATCGTTG 5669
QY 96 lLysGlyLeuMetSerLeuGlyAsnAlaLeuValMetLeuCysValValValLeuMetT 116
Db 5670 AGAAGGGATGATGTCACTTGGAAATGCTCTCGTCATGTTTGTGCGTGTGCTTCTGATGA 5729
QY 116 hrValLeuLeuIleValPheTyrLysTyrLysPheTyrLysLeuIleHisGlyTrpLeuI 136
Db 5730 CAGTTCTGCTGATTTGTTTCTATTAATACAGTTTATTAACGTTTATTAACGTTTATTA 5789
QY 136 leValSerSerPheLeuLeuPheLeuPheThrThrIleTyrValGln----- 152

```



```
Db 11 GAGCAGATGAGGAGAGATGAGGAGCTGACATTTGAAATATATGGCCCAAGCATGTGATC 70
Qy 52 HisLeuPheValProValSerLeuCysMetAlaLeuValPheThrMetAsnThrile 71
Db 71 ATGCTCTTTGTCCTGACTCTCTCATGTTGGTGGTGGTCACTACCATTAAGTCAGTC 130
Qy 72 ThrPheTyrSerGlnAsnAsnGlyArgHisLeuLeuSerHisProPheValArgGluThr 91
Db 131 AGCTTTTATACCGGAGGATGGCAG---CTAATCTATACCCCAATTCACAGAGATACC 187
Qy 92 AspSerIleValGluLeuGlyLeuMetSerLeuGlyAsnAlaLeuValMetLeuCysVal 111
Db 188 GAGACTGTGGCCAGAGAGCCCTGCACTCAATCTGAATGCTGCCATCATGATCATGTC 247
Qy 112 ValValLeuMetThrValLeuLeuIleValPheTyrLysTyrLysPheTyrLysLeuIle 131
Db 248 ATTGTGTGATGACTATCTCTCTGTTGGTCTGTGATAAAATACAGGTGCTATAAGGTCA 307
Qy 132 HisGlyTrpLeuLeuValSerPheLeuLeuPheLeuPheThrThrIleTyrVal 151
Db 308 CATGCTGGCTTATTATATCACTCTATTGTTGCTGTTCTTTTTCATTCATTTACTTGG 367
Qy 152 GlnGluValLeuLysSerPheAspValSerProSerAlaLeuLeuValLeuPheGlyLeu 171
Db 368 GGGGAGTGTAAACCTATAACGTGTGCTGTGCACTACATTACTGTTGCATCTCCTGATC 427
Qy 172 GlyAsnTyrGlyValLeuGlyMetMetCysIleHisTrpLysGlyProLeuArgLeuGln 191
Db 428 TGGAAATTTTGGTGTGGTGGAAATGATTTCCATTCAGTGAAGAGTCCACTTCGACTCCAG 487
Qy 192 GlnPheTyrLeuIleThrMetSerAlaLeuMetAlaLeuValPheIleLysTyrLeuPro 211
Db 488 CAGGCATATCTCATGATTGATGATTAGTCCCTCATGGCCCTGTTGTTATCAAGTACCTCCCT 547
Qy 212 GluTrpThrValTrpPheValLeuPheValIleSerValTrpAspLeuValAlaValLeu 231
Db 548 GAATGCACTGCTGGCTCATCTTGGCTGTGATTCAGTATATGATTTAGTGGCTGTTTG 607
Qy 232 ThrProLysGlyProLeuArgTyrLeuValGluThrAlaGlnGluArgAsnGluProIle 251
Db 608 TGTCCGAAAGGTCCTACTTGGTATGCTGGTGAACAGCTCAGAGAGAAATGAAAGCCTT 667
Qy 252 PheProAlaLeuIleTyrSerGlyValIleTyrProTyrValLeuValThrAlaVal 271
Db 668 TTCCAGCTCTCATTTACTCTCAATATGGTGTGG---TTGGTGAATATGGCA 718
Qy 272 GluAsnThrThrAspProArgGluProThrSerSerAspSerAsnThrSerThrAlaPhe 291
Db 719 GAAGGAGACCCGGAAGCTCAAGAGGAGATATCAAAATTTCC----- 760
Qy 292 ProGlyGluAlaSerCysSerSerGluThrProLysArgProLysValLysArgIlePro 311
Db 761 -----AAGTATAATCAGAAAGACAGAAAGG----- 787
Qy 312 GlnLysValGlnIleLysSerAsnThrThrAlaSerThrThrGlnAsnSerGlyValArg 331
Db 788 -----GAGTCAACAGACATGTTGCAGAAATGATGCGCGGTTCAGT 832
Qy 332 ValGlu-----ArgGluLeuAlaAlaGlu 339
Db 833 GAGGAATGGGAAGCCAGAGGACAGTCATCTAGGCGCTCATCGCTCTACACCTCAGTCA 892
Qy 340 ArgProThrValGlnAspAlaAsn-----PheHisArgHisGluGluGluArg 356
Db 893 CGAGCTGCTGTCCAGGAACCTTTCCAGCAGTATCTCGCTGTGTGAAGACCCAGGAAAGG 952
Qy 357 GlyValLysLeuGlyLeuGlyAspPheIlePheTyrSerValLeuLeuGlyLysAlaSer 376
Db 953 GGAGTAAACCTTGGATTTGGAGATTTTCAATTTCTACAGTGTCTGTTGGTGAAGCCCTCA 1012
Qy 377 SerTyrPhe-----AspTrpAsnThrThrIleAlaCysTyrValAlaIleLeuIleGly 394
Db 1013 GCAACAGCAGCTGGGAGACTGGAAACAACCATATGCTGTTTCGTAGCCATATTATTTGGT 1072
```

```
Qy 395 LeuCysPheThrLeuValLeuLeuAlaValPheLysArgAlaLeuProAlaLeuGlnPh 414
Db 1073 TTGTGCTTTACATTATTACTCTTGCCATTTTCAAGAAAGCATTTGCCAGCTCTTCCAATC 1132
Qy 414 eProPheSerProAspSerPhePheThrPheValProAlaGlySerSerProHisLeuLe 434
Db 1133 TCCA---TCACCTTTGGGCTTGTGTTTCTACTTTGCCACAGATTTATCTGTACAGCCTTT 1189
Qy 434 uHisLysSerLeuLysSerValTyrTyrIleAsnSerLeuPhe-----LeuProPh 451
Db 1190 ATGGACCAATTAGCATTCATCAATTTTATATCTAGCATATTTGCCGTAGAAATCCCATG 1249
Qy 451 eLeuCysIleIle 455
Db 1250 GATGTTTCTCTT 1262

RESULT 7
LOCUS AR095621 2764 bp DNA linear PAT 08-SEP-2000
DEFINITION Sequence 9 from patent US 6004794.
ACCESSION AR095621
VERSION AR095621.1 GI:10023656
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2764)
AUTHORS Karran,E.Howard., Clinkenbeard,H.Elizabeth., Browne,M.Joseph.,
Southan,C.David., Creasy,C.Lee. and Iivi,G.Pietro.
TITLE Human serine protease
JOURNAL Patent: US 6004794-A 9 21-DEC-1999;
FEATURES Location/Qualifiers
source
1..2764
/organism="unknown"

BASE COUNT 715 a 623 c 652 g 773 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 6,61e-78 Length: 2764
Score: 993.50 Matches: 216
Percent Similarity: 65.84% Conservative: 77
Best Local Similarity: 48.54% Mismatches: 104
Query Match: 41.78% Indels: 148
DB: 6 Gaps: 9

US-09-043-944-1 (1-461) x AR095621 (1-2764)
Qy 32 GlnGluAspGluAsnValValGluGluAlaGluLeuLysTyrGlyAlaSerHisValIle 51
Db 438 GAGCAAGATGAGGAGAGAGATGAGGAGCTGACATTTGAAATATGGCCCAAGCATGTGATC 497
Qy 52 HisLeuPheValProValSerLeuCysMetAlaLeuValValPheThrMetAsnThrIle 71
Db 498 ATGCTCTTTTGTCCCTGTGACTCTCTGCATGTTGGTGGTGGTGGTGGTGGTGGTGGTGGT 557
Qy 72 ThrPheTyrSerGlnAsnAsnGlyArgHisLeuLeuSerHisProPheValArgGluThr 91
Db 558 AGCTTTTATACCGGAGGATGGGAG---CTAATCTATACCCCAATTCACAGAAATACC 614
Qy 92 AspSerIleValGluLysGlyLeuMetSerLeuGlyAsnAlaLeuValMetLeuCysVal 111
Db 615 GAGACTGTGGCCAGAGAGCCCTGCACTCAATCTGAAATGCTGCTGCTGCTGCTGCTGCTGCT 674
Qy 112 ValValLeuMetThrValLeuLeuIleValPheTyrLysTyrLysPheTyrLysLeuIle 131
Db 675 ATTGTGTGATGACTATCTCTCTGTTGGTCTGTGATAAAATACAGGTGCTATAAGGTCA 734
Qy 132 HisGlyTrpLeuLeuValSerSerPheLeuLeuLeuPheLeuPheThrThrIleTyrVal 151
Db 735 CATGCTGGCTTATTATATCACTCTATGTTGCTGTTCTTTTTCATTCATTTACTTGTG 794
Qy 152 GlnGluValLeuLysSerPheAspValSerProSerAlaLeuLeuValPheGlyLeu 171
```


Qy 192 GlnPheTyrLeuIleThrMetSerAlaLeuMetAlaValPheLeuTyrLeuPro 211
Db 915 CAGCATATCTCATATGATAGTCCCTATGCGCTGTTTATCAAGTACCTCCCT 974
Qy 212 GluTrpThrValTrpPheValLeuPheValIleSerValTrpAspLeuValAlaValLeu 231
Db 975 GAATGGAGCTGGTGCATCTTGGCTGTGATTTTCAGTATATGATTTAGTGGCTGTTTG 1034
Qy 232 ThrProLysGlyProLeuArgTyrLeuValGluThrAlaGlnGluArgAsnGluProIle 251
Db 1035 TGTCCGAAGGCTCCATCTCGTATCTGTTGAAACAGCTCAGGAGAGAAATGAACGCTT 1094
Qy 252 PheProAlaLeuIleTyrSerSerGlyValIleTyrProTyrValLeuValAlaVal 271
Db 1095 TTTCAGCTCTCTATTACTCTCAACAATGGTGG-----TTGGTGAATATGGCA 1145
Qy 272 GluAsnThrThrAspProArgGluProThrSerSerAspSerAsnThrSerThrAlaPhe 291
Db 1146 GAAGGAGACCCGGAAGCTCAAGGAGAGATATCCAAATATCC----- 1187
Qy 292 ProGlyGluAlaSerCysSerSerGluThrProLysArgProLysValLysArgIlePro 311
Db 1188 -----AGTATAATGCGAAGACACAGAAAG----- 1214
Qy 312 GlnLysValGlnIleGluSerAsnThrThrAlaSerThrThrGlnAsnSerGlyValArg 331
Db 1215 -----GAGTCAACAAGACACTTTCAGAGAATGATGATGCGGGTTCAGT 1259
Qy 332 ValGlu-----ArgGluLeuAlaAlaGlu 339
Db 1260 GAGGATGGGAAGCCGAGGAGGAGCAGTCATCTAGGGCTCTACCTCTACACCTGAGTCA 1319
Qy 340 ArgProThrValGlnAspAlaAsn-----PheHisArgHisGluGluGluArg 356
Db 1320 CGAGCTGCTGTCAGGAATTTCCAGCAGTATCTCGCTGCTGGAAGACCCAGAGGAAGG 1379
Qy 357 GlyValLysLeuGlyLeuGlyAspPheIlePheTyrSerValLeuLeuGlyLysAlaSer 376
Db 1380 GGAGTAAACACTTGAATGGGAGATTTCAATTTCTACAGTGTCTGTGGTAAAGCCTCA 1439
Qy 377 SerTyrPhe-----AspTrpAsnThrThrIleAlaCysTyrValAlaIleLeuIleGly 394
Db 1440 GCAACAGCAGTGGAGACTGGACACACACATAGCTTTCGTAGCCATATTAATGGT 1499
Qy 395 LeuCysPheThrLeuValLeuAlaValPheLysArgAlaLeuProAlaLeu-GlnPh 414
Db 1500 TTGTGCTTACATATTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1559
Qy 414 eProPheSerProAspSerPhePheThrPheValProAlaGlySerSerProHisLeuLe 434
Db 1560 TCCA---TCACCTTGGGCTGTTTCTACCTTGGCCACAGATATCTGTACAGCCTTTT 1616
Qy 434 uHisLysSerLeuLysSerValTyrTyrIleAsnSerLeuPhe-----LeuProPh 451
Db 1617 ATGGACCAATTAGCATTCATCAATTTTATATCTAGCATATTTGGGTTAGAAATCCCATG 1676
Qy 451 eLeuCysIleIle 455
Db 1677 GATGTTTCTCTCT 1689
RESULT 9
LOCUS AX244648
DEFINITION Sequence 4 from Patent WO0167097.
ACCESSION AX244648
VERSION AX244648.1 GI:15859540
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1

AUTHORS Hale, R.S., Rowley, A.G. and Blackstock, W.G.

TITLE Novel assay

JOURNAL Patent: WO 0167097-A 4 13-SEP-2001;

GLAXO GROUP LIMITED (GB)

FEATURES

source Location/Qualifiers

1. 2765

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

249..1652

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC88578.1"

/translation="TELPAPLSYFONQMSGDHLSNTVRSQNDNRQEHNDRRSL

GHEPLSNRPGQNSPOVVEODEEDELTKYGAHVLMFLPVPVTLQWVVVATIKS

VSFTRKQQLIYTFETEDTVGQORALHSILNRAIMSLIVMTLLVLYKRCYK

VIHMLIISLLELFFSYFIVGKFNVAVDITVALLIWNVGVGMISIHKGP

LRLQAYLIMISALMALVFIKYPENTAWLILAVISVDLVLCPKPLRLMLVETAG

ERNETLPALYSSTWMLVNMADEPAQRVSKNSNAESTERESQDTVAENDQ

GFSEBWAQRDSHLGPHRSTPESRAAVQELSSILAGEDPBERGVKLGIDPFIYSVL

VGKASATAGDMWNTIACFVAILGLCLTLALLAIFKKALPALPISITFGLVYFATD

YLVSQSPMDQLAFHQFYI"

BASE COUNT 715 a 624 c 652 g 773 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 6-62e-78 Length: 2765

Score: 993.50 Matches: 216

Percent Similarity: 65.84% Conservative: 77

Best Local Similarity: 48.54% Mismatches: 104

Query Match: 41.78% Indels: 48

DB: 6 Gaps: 9

US-09-043-944-1 (1-461) x AX244648 (1-2765)

Qy 32 GlnGluAspGluAsnValValGluGluAlaGluLeuLysTyrGlyAlaSerHisValIle 51

Db 438 GAGCAAGATGAGGAGAGAGATGAGGAGCTGACATGAAATATGGCGCCCAAGCATGTGATC 497

Qy 52 HisLeuPheValProValSerLeuCysMetAlaLeuValValPheThrMetAsnThrIle 71

Db 498 ATGCTCTTTGTCCTGACTCTCTGATGTTGGTGGTCTGCTGCTACCATTAAGTCAGTC 557

Qy 72 ThrPheTyrSerGlnAsnAsnGlyArgHisLeuLeuSerHisPhePheValArgGluThr 91

Db 558 AGCTTTTATACCGGAGGATGGGAG---CTAATCTATACCCCATTCACAGAAGATACC 614

Qy 92 AspSerIleValGluLysGlyLeuMetSerLeuGlyAsnAlaLeuValMetLeuCysVal 111

Db 615 GAGACTGTGGCGAGAGACCTTGCACTCAATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 674

Qy 112 ValValLeuMetThrValLeuLeuIleValPheTyrLysTyrLysPheTyrLysLeuIle 131

Db 675 ATGTGTCATGACTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 734

Qy 132 HisGlyTrpLeuIleValSerSerPheLeuPheLeuPheThrThrIleTyrVal 151

Db 735 CAGGCTGGCTTATATATCATCTATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 794

Qy 152 GlnGluValLeuLysSerPheAspValSerProSerAlaLeuLeuValPheGlyLeu 171

Db 795 GCGGAAGTGTATAAACCTATAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 854

Qy 172 GlyAsnTyrGlyValLeuGlyMetMetCysIleHisTrpLysGlyProLeuArgLeuGln 191

Db 855 TGGAAATTTTGGTGGTGGGAATGATTTCCATTCATCGGAAAGGTCCTTCGACTCCAG 914

Qy 192 GlnPheTyrLeuIleThrMetSerAlaLeuValPheIleLysTyrLeuPro 211

Db 915 CAGGCATATCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 974

Qy 212 GluTrpThrValTrpPheValLeuPheValIleSerValTrpAspLeuValAlaValLeu 231

Db 1380 GGAGTAAACTGGATTGGGAGATTTCATTTTCTACAGTGTCTCTGGTGGTAAAGCCTCA 1439
 QY 377 SerTyrPhe-----AspTrpAsnThrThrIleAlaCysTyrValAlaIleLeuIleGly 394
 Db 1440 GCAACAGCCAGTGGAGCTGGAAACAACCATAGCTGTTCGTAGCCATATTAATTTGGT 1499
 QY 395 LeuCysPheThrLeuValLeuAlaValPheLysArgAlaLeuProAlaLeu-GlnPh 414
 Db 1500 TTGTGCTTACATATTACTCTTGGCATTTTCAAGAAAGCATGGCCAGCTCTTCCAATC 1559
 QY 414 eProPheSerProAspSerPhePheThrPheValProAlaGlySerSerProHisLeuLe 434
 Db 1560 TCCA---TCACCTTTGGCTGTCTTCTACTTTGCCACAGATATCTTTGTACAGCTTTT 1616
 QY 434 uHisLysSerLeuLysSerValTyrTyrIleAsnSerLeuPhe-----LeuProPh 451
 Db 1617 ATGGACCAATAGCATTCATCAATTTTATATCTAGCATATTTGCGGTAGAAATCCCATG 1676
 QY 451 eLeuCysIleIle 455
 Db 1677 GATGTTCTCTCTT 1689

RESULT 15
 AR256768
 LOCUS AR256768 2791 bp mRNA linear PAT 20-DEC-2002
 DEFINITION Sequence 133 from patent US 6485911.
 ACCESSION AR256768
 VERSION AR256768.1 GI:27306376
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 2791)
 AUTHORS St. George-Hyslop,P.H., Rommens,J.M. and Fraser,P.E.
 TITLE Methods for determining risk of developing alzheimer's disease by
 detecting mutations in the presenilin 2 (PS-2) gene
 JOURNAL Patent: US 6485911-A 133 26-NOV-2002;
 FEATURES Location/Qualifiers
 source 1..2791
 BASE COUNT 742 a 624 c 652 g 773 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6,7e-78 Length: 2791
 Score: 993.50 Matches: 216
 Percent Similarity: 65.84% Conservative: 77
 Best Local Similarity: 48.54% Mismatches: 104
 Query Match: 41.78% Indels: 48
 DB: 6 Gaps: 9

US-09-043-944-1 (1-461) x AR256768 (1-2791)

QY 32 GlnGluAspGluAsnValValGluGluAlaGluLeuLysTyrGlyAlaSerHisValIle 51
 Db 438 GAGCAAGATGAGGAAGATGAGGAGCTGACATTCGAATATGGCCCAAGCATGTGATC 497
 QY 52 HisLeuPheValProValSerLeuCysMetAlaLeuValPheThrMetAsnThrIle 71
 Db 498 ATGCTCTTTGTCCTGTGACTCTCTGCATGTGTGTGCTGCTGCTGCTGCTGCTGCTGCT 557
 QY 72 ThrPheTyrSerGluAsnGlyArgHisLeuLeuSerHisProPheValArgGluThr 91
 Db 558 AGCTTTTATACCGGAAGATGGGAG---CTAATCTATACCCCATTCACAGAGATACC 614
 QY 92 AspSerIleValGluLysGlyLeuMetSerLeuGlyAsnAlaLeuValMetLeuCysVal 111
 Db 615 GAGACTGTGGCCAGAGACCCCTGCACTCAATCTGAATCTGCCCATCATGATCAGTGTC 674
 QY 112 ValValLeuMetThrValLeuLeuIleValPheTyrIleTyrLysPheTyrLysLeuIle 131
 Db 675 ATTGTGTGATGACTATCTCTGCTGGTGTCTGTATATAATACAGGGTGTCTATAAGGTCATC 734

QY 132 HisGlyTyrIleValSerSerPheLeuLeuLeuPheLeuPheThrThrIleTyrVal 151
 Db 735 CATGCTGGCTTATTATATCATCTATTGTTGCTCTTTTTCATTTCACTTACTTG 794
 QY 152 GlnGluValLeuLysSerPheAspValSerProSerAlaLeuValLeuPheGlyLeu 171
 Db 795 GGGGAAGTGTAAACCTATACGTGCTGTGGACTACATTACTGTGTGACTCTGATC 854
 QY 172 GlyAsnTyrGlyValLeuGlyMetMetCysIleHisTyrLysGlyProLeuArgLeuGln 191
 Db 855 TGGAAATTTTGGTGTGGTGGATGATTTCCATTTCACTGGAAAGGTCCACTTCGACTCCAG 914
 QY 192 GlnPheTyrIleThrMetSerAlaLeuMetAlaLeuValPheIleLysTyrLeuPro 211
 Db 915 CAGGCATATCTCATTTAGTATGATGCTCATGGCCCTGGTGTATATCAAGTACCTCCCT 974
 QY 212 GluTyrThrValTyrPheValLeuPheValIleSerValTyrAspLeuValAlaValLeu 231
 Db 975 GAATGGACTGCGTGGCTCATCTTGGCTGTGATTTTCAGTATATGATTTAGTGGCTGTTTG 1034
 QY 232 ThrProLysGlyProLeuArgTyrLeuValGluThrAlaGlnGluArgAsnGluProIle 251
 Db 1035 TGTCCGAAAGGTCCACTTCGTATGCTGTGTTGAAACAGCTCAGCAGAGAATGAACGCTT 1094
 QY 252 PheProAlaLeuIleTyrSerSerGlyValIleTyrProTyrValLeuValThrAlaVal 271
 Db 1095 TTTCAGCTCTCATTTTACTCTCAACATGGTGTGG-----TTGGTGAATATGGCA 1145
 QY 272 GluAsnThrThrAspProArgGluProThrSerSerAspSerAsnThrSerThrAlaPhe 291
 Db 1146 GAAGGAGCCCGAAGCTCAAAGGAGATATCCAAAATTC-----1187
 QY 292 ProGlyGluAlaSerCysSerSerGluThrProLysArgProLysValLysArgIlePro 311
 Db 1188 -----AAGTATAATCAGAAAGCAGACAGAAAGG-----1214
 QY 312 GlnLysValGlnIleGluSerAsnThrThrAlaSerThrThrGlnAsnSerGlyValArg 331
 Db 1215 -----GAGTCACAAGACACTGTTGCAGAAATGATGTCGGGTTCAGT 1259
 QY 332 ValGlu-----ArgGluLeuAlaAlaGlu 339
 Db 1260 GAGGAATGGGAAGCCAGAGGACAGTCATCTAGGCCCTCATCGCTCTACCCCTGAGTCA 1319
 QY 340 ArgProThrValGlnAspAlaAsn-----PheHisArgHisGluGluGluArg 356
 Db 1320 CGAGCTGCTGTCCAGGAACCTTCCAGCAGTATCTCGCTGGTGAAGACCCAGAGAAAGG 1379
 QY 357 GlyValLysLeuGlyLeuGlyAspPheIlePheTyrSerValLeuLeuGlyLysAlaSer 376
 Db 1380 GGAGTAAACCTTGGATTGGGAGATTTCAATTTCTACAGTGTCTGTTGGTAAAGCCTCA 1439
 QY 377 SerTyrPhe-----AspTrpAsnThrThrIleAlaCysTyrValAlaIleLeuIleGly 394
 Db 1440 GCAACAGCCAGTGGAGACTGGAAACAACCATAGCTGTTCGTAGCCATATTAATTTGGT 1499
 QY 395 LeuCysPheThrLeuValLeuAlaValPheLysArgAlaLeuProAlaLeu-GlnPh 414
 Db 1500 TTGTGCTTACATATTACTCTTGGCATTTTCAAGAAAGCATTTGCCAGCTCTTCCAATC 1559
 QY 414 eProPheSerProAspSerPhePheThrPheValProAlaGlySerSerProHisLeuLe 434
 Db 1560 TCCA---TCACCTTTGGGCTTGTCTTCTACTTTGCCACAGATATCTTTGTACAGCTTTT 1616
 QY 434 uHisLysSerLeuLysSerValTyrTyrIleAsnSerLeuPhe-----LeuProPh 451
 Db 1617 ATGGACCAATAGCATTCATCAATTTTATATCTAGCATATTTGCGGTAGAAATCCCATG 1676
 QY 451 eLeuCysIleIle 455
 Db 1677 GATGTTCTCTCTT 1689

Search completed: December 11, 2003, 07:41:27
Job time : 4266 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 11, 2003, 06:24:06 ; Search time 2279 Seconds
(without alignments)
4916.351 Million cell updates/sec

Title: US-09-043-944-1
Perfect score: 2378
Sequence: 1 MPSTRRQEGGGAETHV.....YYINSLFLPFLCINFSIIS 461

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US09043944/runat_11122003_062400_3685/app_query.fasta_1.647
-DB=EST -QFMT=Fastap -SUPFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09043944 @CGN 1_3349 @runat_11122003_062400_3685 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hcc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	Length	DB	ID	Description
1	986	41.5	589	14	CB385197	CB385197	OSTF019E1
2	944	33.7	562	14	CB391622	CB391622	OSTF155F5
C 3	942	33.6	558	14	CB404201	CB404201	OSTR019E1
C 4	834	35.1	494	14	CB395382	CB395382	OSTR155F5
5	788	33.1	477	12	BI174328	BI174328	OSTF019E1
C 6	734.5	30.9	884	14	CD513489	CD513489	AGENCOURT
7	734	30.9	964	13	EX426340	EX426340	AGENCOURT
8	724.5	30.5	1101	12	BM548100	BM548100	AGENCOURT
9	713.5	30.0	1998	11	AK014706	AK014706	Mus muscu
10	702.5	29.5	661	9	AI675803	AI675803	wb97a12.x
11	699	29.4	697	13	BU614355	BU614355	UI-M-EVO-
12	698.5	29.4	703	9	AI925372	AI925372	wn53406.x
C 13	688.5	29.0	974	13	EX426339	EX426339	EX426339
14	681.5	28.7	680	9	AI831581	AI831581	wj39404.x
15	671	28.2	881	14	CD358381	CD358381	AGENCOURT
16	665.5	28.0	797	13	BU847845	BU847845	AGENCOURT
C 17	645	27.1	386	12	BI175258	BI175258	OSTR019E1
18	644	27.1	943	14	CD512951	CD512951	AGENCOURT
19	631	26.5	757	10	AW850572	AW850572	IL3-CT021
20	631	26.5	1112	12	BM476222	BM476222	AGENCOURT
21	628	26.4	735	10	AW850636	AW850636	IL3-CT021
22	627.5	26.4	705	9	AW604230	AW604230	IL3-CT021
23	627	26.4	733	10	AW850631	AW850631	IL3-CT021
24	625	26.3	671	10	AW850640	AW850640	IL3-CT021
25	625	26.3	714	9	AW604310	AW604310	IL3-CT021
26	623.5	26.2	874	14	CA477482	CA477482	AGENCOURT
27	623	26.2	1201	9	AL553513	AL553513	AL553513
28	621	26.1	375	14	CA7805	CA7805	CA7805
29	620	26.1	360	9	AV190680	AV190680	AV190680
30	620	26.1	640	9	AW604295	AW604295	IL3-CT021
31	614	25.8	980	14	BY714205	BY714205	BY714205
32	612	25.7	1201	9	AL553084	AL553084	IL3-CT021
33	611.5	25.7	601	9	AW604307	AW604307	IL3-CT021
34	610.5	25.7	613	9	AW367465	AW367465	MR0-HT016
35	609.5	25.6	581	10	AW850712	AW850712	IL3-CT021
36	609	25.6	374	14	C66636	C66636	C66636
37	609	25.6	704	10	EG633935	EG633935	AT29922.5
38	608.5	25.6	587	9	AA993681	AA993681	ot97b02.s
39	607.5	25.5	591	10	AW850561	AW850561	IL3-CT021
40	607.5	25.5	592	9	AW604313	AW604313	IL3-CT021
41	607.5	25.5	599	9	AW604231	AW604231	IL3-CT021
42	606.5	25.5	937	14	CA496597	CA496597	AGENCOURT
43	602.5	25.3	944	12	BI453842	BI453842	603174713
44	600.5	25.3	593	10	AW850573	AW850573	IL3-CT021
C 45	600.5	25.3	737	10	AW850632	AW850632	IL3-CT021

ALIGNMENTS

RESULT 1
CB385197
LOCUS
DEFINITION
CB385197
ACCESSION
VERSION
CB385197.1
KEYWORDS
EST.
SOURCE
Caenorhabditis elegans
ORGANISM
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
1 (bases 1 to 589)
Reboul,J., Vaglio,P., Rual,J.F., Lamesch,P., Martinez,M., Armstrong
CB385197 589 bp mRNA linear EST 15-MAY-2003
OSTF019E11_2 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
CB385197.1 GI:30726907

Query Match: 39.70% Indels: 0
 DB: 14 Gaps: 0
 US-09-043-944-1 (1-461) x CB391622 (1-562)

Qy 8 GlnGluGlyGlyGlyAlaHepAlaGluThrHisThrValTyGlyThrAsnLeuLeuThr 27
 Db 3 CAGGAGGCGGAGGTGCAGATCGGAAACACATACCGTTTACGGTACAAATCTGATAACA 62
 Qy 28 AsnArgAsnSerGlnGluAspGluAsnValValGluGluAlaGluLeuLeuTyGlyAla 47
 Db 63 AATCGGAATAGCAAGAACCAAAATGTTGGAGAGCGGAGCTGAAATACGGAGCA 122
 Qy 48 SerHisValIleHisLeuPheValProValSerLeuLeuCysMetAlaLeuValValPheThr 67
 Db 123 TCTCACGTTATTCACTATTGTCGGGTGTCACTATGTCATGGCTCTGGTGTGTTTACG 182
 Qy 68 MetAsnThrIleThrPheTySerGlnAsnAsnGlyArgHisLeuLeuSerHisProPhe 87
 Db 183 ATGAACACGATTACGTTTATATAGTCAAAACAATGGAAGCAATTTACTATACACTCTCTTTT 242
 Qy 88 ValArgGluThrAspSerIleValGluLeuGlyLeuMetSerLeuGlyValAsnAlaLeuVal 107
 Db 243 GTCCGGGAACACAGACAGTATCGTTGAGAGGATGATGTCATCTGGAAATGCTCTGTC 302
 Qy 108 MetLeuCysValValValLeuMetThrValLeuLeuLeuIleValPheTyLeuTyLeuPhe 127
 Db 303 ATGTTGTGGGTGCTGCTGATGACAGTCTGCTGATTTCTTCTATAAATACAGTTT 362
 Qy 128 TyrLeuLeuIleHisGlyTrpLeuIleValSerSerPheLeuLeuLeuPheLeuPheThr 147
 Db 363 TATAAGCTTATTCATGAGTGGCTTATGTCACAGTTTCTTCTTCTTCTTCTTCTTCT 422
 Qy 148 ThrIleTyValGlnGluValLeuLeuLeuSerPheAspValSerProSerAlaLeuLeuVal 167
 Db 423 ACAATCTATGTGCAAGAGTCTGAAAGTTTCGATGTCTCCACGCGCACTATTGTT 482
 Qy 168 LeuPheGlyLeuGlyAsnTyGlyValLeuGlyMetMetCysIleHisTrpLeuGlyPro 187
 Db 483 TTGTTGGACTGGTAACTATGAGGTCTCGGAATGATGTATATACATTGGAAGGTCCA 542
 Qy 188 LeuArgLeuGlnGlnPhe 193
 Db 543 TTGCGTCTGCAAGATTC 560

RESULT 3
 CB404201/c
 LOCUS
 DEFINITION OSTRO19E11_2 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
 ACCESSION CB404201
 VERSION CB404201.1 GI:30745928
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
 1 (bases 1 to 558)
 Rebol,J., Vaglio,P., Rual,J.F., Lamesch,P., Martinez,M., Armstrong
 C.M., Li,S., Jacotot,L., Bertin,N., Janky,R., Moore,T., Hudson
 J.R., Hartley,J.L., Brasch,M.A., Vandenhaute,J., Boulton,S.,
 Endress,G.A., Jenna,S., Chevet,E., Papasotiropoulos,V., Tollas,P.P.,
 Pracek,J., Snyder,M., Huang,R., Chance,M.R., Lee,H.,
 Doucette-Stamm,L., Hill,D.E. and Vidal,M.
 C. elegans ORFome version 1.1: experimental verification of the
 genome annotation and resource for proteome-scale protein
 expression

TITLE
 Nat. Genet., (2003) In press
 Contact: Vidal M
 Marc Vidal Laboratory
 Dana Farber Cancer Institute
 1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
 Tel: 617 632 5180
 Fax: 617 632 5739

JOURNAL
 COMMENT

Email: Marc Vidal@fci.harvard.edu
 Sequence tag of Gateway entry clones. The primers used were
 designed on the predicted protein encoding ORF. C. elegans ORFome
 cloning project : Contact david_hill@fci.harvard.edu or
 marc_vidal@fci.harvard.edu
 POLYA=No.

FEATURES
 Source
 Location/Qualifiers
 1..558
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /sex="Hermaphrodite and male"
 /tissue_type="whole animal"
 /dev_stage="mixed stage"
 /clone_lib="AD-wrmcDNA"
 /notes="The AD-wrmcDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC86"
 BASE COUNT 153 a 118 c 149 g 138 t
 ORIGIN

Alignment Scores:
 Pred. No.: 8,43e-84 Length: 558
 Score: 942.00 Matches: 185
 Percent Similarity: 99.46% Conservative: 0
 Best Local Similarity: 99.46% Mismatches: 0
 Query Match: 39.61% Indels: 1
 DB: 14 Gaps: 0

US-09-043-944-1 (1-461) x CB404201 (1-558)

Qy 249 GluProIlePheProAlaLeuIleTySerSerGlyValIleTyProTyValLeuVal 268
 Db 557 GAGCCATTTTCCCGCGCTGATTTATTCCTGAGTCACTATCCCTACGTTCTGTT 498
 Qy 269 ThrAlaValGluAsnThrThrAspProArgGluProThrSerSerAspSerAsnThrSer 288
 Db 497 ACTGCACTTGAACACACAGACAGACCCCGTGAACCGACGTCGTCAGACTCAAACTCTCT 438
 Qy 289 ThrAlaPheProGlyGluAlaSerCysSerSerGluThrProLysArgProLysValLys 308
 Db 437 ACAGCTTTTCTGGAGAGGCGAGTTGTTCACTGAAACGCCAAACGCCCAAAAGTGAAA 378
 Qy 309 ArgIleProGlnLysValGlnIleGluSerAsnThrThrAlaSerThrThrGlnAsnSer 328
 Db 377 CGAATTCCTCAAAAAGTGCAAATCGAATCGAATACGATTCAGCTTCAACGACACAAACTCT 318
 Qy 329 GlyValArgValGluArgGluLeuAlaAlaGluArgProThrValGlnAspAlaAsnPhe 348
 Db 317 GGAGTAAGGGTGGAAACGGGAGCTAGCTGCTGAGAGACCAACTGTACAAGACGCCAATTTT 258
 Qy 349 HisArgHisGluGluGluArgGlyValLysLeuGlyLeuGlyLeuGlyAspPheIlePheTy 368
 Db 257 CACAGGCACGACGAGGAGAGAGAGGTGAAACTTGGTCTGGCGGACTTCATTTCTAC 198
 Qy 369 SerValLeuLeuGlyLysAlaSerSerTyPheAspTrpAsnThrThrIleAlaCysTy 388
 Db 197 TCTGTTCTCTCGCAAGGCTTCATCGTACTTTTGACTTGAACACGACTATCGTCTGTTAT 138
 Qy 389 ValAlaIleLeuIleGlyLeuCysPheThrLeuValLeuAlaValPheLysArgAla 408
 Db 137 GTGGCCATTTTATTCGGTCTCTGCTTCACCTTGTCTGCTGCCCTCTTCAACGAGCA 78
 Qy 409 LeuProAlaLeu-GlnPheProPheSerProAspSerPhePheThrPheValProAlaG 428
 Db 77 CTCCGGGCTCTGCCAATTTCCATTTTCTCGGACTCATTTTGTACCTTTGTACCCGCTG 18
 Qy 428 YSerSerProHisLeu 433
 Db 17 ATCATCACCCCATTTG 2

```

RESULT 4
CB395382/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE

JOURNAL
COMMENT

FEATURES
    source
        1..494
            /organism="Caenorhabditis elegans"
            /mol_type="mRNA"
            /strain="N2"
            /db_xref="taxon:6239"
            /sex="Hermaphrodite and male"
            /tissue_type="whole animal"
            /dev_stage="mixed stage"
            /clone_lib="AD-wrmcDNA"
            /note="The AD-wrmcDNA library was generated with poly(A)+
            RNA isolated from both hermaphrodite and male N2 worms of
            all larval stages, embryos, adults and dauers and the
            subsequent generation of cDNAs by poly(A) priming. The
            cDNAs were cloned into pPC86"
            147 a 100 c 123 g 124 t
BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.:      4,44e-73      Length:      494
Score:          834.00      Matches:    164
Percent Similarity: 99.30%      Conservat:  0
Best Local Similarity: 99.30%      Mismatches:  0
Query Match:    35.07%      Indels:     1
DB:             14          Gaps:         0

US-09-043-944-1 (1-461) x CB395382 (1-494)

QY 289 ThrAlaPheProGlyGluAlaSerCysSerSerGluThrProLysArgProLysValLys 308
DB 494 ACAGCTTTTCTCGAGGAGCGAGTGTTCATCTGAAACGCCAAACGGCCAAAGTGAA 435
QY 309 ArgGleProGlyLysValGlnIleGluSerAsnThrThrAlaSerThrThrGlnAsnSer 328
DB 434 CGAATCTCTCAAAAGTGCAAATCGAATCGAATCTACAGCTTCACAGCACACAACTCT 375
QY 329 GlyValArgValGluArgGluLeuAlaAlaGluArgProThrValGlnAspAlaAsnPhe 348

```

```

Db 374 GGAGTAAGGCTGGACCGGAGCTAGCTGTGAGAGACCAACTGTACAAAGACGCCAATTTT 315
QY 349 HisArgHisGluGluGluArgGlyValLysLeuGlyLeuGlyAspPheIlePheTyr 368
Db 314 CACAGCACGACGAGAGAGAGAGAGAGTGTGAAACTTGGTCTGGCGGACTTCATTTTCTAC 255
QY 369 SerValLeuLeuGlyLysAlaSerSerTyrPheAspTrpAsnThrThrIleAlaCysTyr 388
Db 254 TCTGTTCTCTCGCGAAGGCTTCATCGTACTTGTGACTGGAAACAGCACTATCGCTTGTAT 195
QY 389 ValAlaIleLeuIleGlyLeuCysPheThrLeuValLeuAlaValPheLysArgAla 408
Db 194 GTGGCCATCTTATCCGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 135
QY 409 LeuProAlaLeu-GlnPheProPheSerProAspSerPheThrPheValProAlaGln 428
Db 134 CTCCTGGCTCTGCCAATTTCCATTTTCTCGGACTCATTTTCTTTTGTATCCCGCTGG 75
QY 428 ySerSerProHisLeuLeuHisLysSerLysSerValTyrTyrIleAsnSerLeuPh 448
Db 74 ATCATCACCCTTTTGTATACAAAGTCTCTCAAAAGTGTATTATATTATTAATTTCTCTGT 15
QY 448 eLeuProPheLeu 452
Db 14 TTGCGCATTTCTT 2

```

```

RESULT 5
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
PUBMED
COMMENT

Contact: Reboul J, Vaglio P
Marc Vidal Laboratory
Dana Farber Cancer Institute
44 Binney Street, Boston, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 2425
Email: Jerome.Reboul@dfci.harvard.edu
Sequence tag of Gateway entry clones. The primers used were
designed on the predicted protein encoding ORF. C. elegans ORFeome
cloning project. Contact jerome_reboul@dfci.harvard.edu or
philippe_vaglio@dfci.harvard.edu
POLYA=No.

```

```

FEATURES
    source
        1..477
            /organism="Caenorhabditis elegans"
            /mol_type="mRNA"
            /strain="N2"
            /db_xref="taxon:6239"
            /sex="Hermaphrodite and male"
            /tissue_type="whole animal"
            /dev_stage="mixed stage"
            /clone_lib="AD-wrmcDNA"
            /note="The AD-wrmcDNA library was generated with poly(A)+
            RNA isolated from both hermaphrodite and male N2 worms of
            all larval stages, embryos, adults and dauers and the

```

subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPC86"

```

BASE COUNT      131 a      87 c      109 g      150 t
ORIGIN
Alignment Scores:
Pred. No.:      1,68e-68      Length:      477
Score:          788.00      Matches:      156
Percent Similarity: 98.73%      Conservative: 0
Best Local Similarity: 98.73%      Mismatches: 2
Query Match:      33.14%      Indels:      0
DB:              12      Gaps:      0

US-09-043-944-1 (1-461) x B1174328 (1-477)
QY      8 GInGluGlyGlyGlyAlaAspAlaGluThrHisThrValTyrGlyThrAsnLeuLeuPheThr 27
DB      3 CAGAGGGCGGAGGTCAGATCGGAAACACATACCGTTTACGGTACAAATCTGATACACA 62
QY      28 AsnArgAsnSerGlnGluAspGluAsnValValGluGluAlaGluLeuLeuTyrGlyAla 47
DB      63 AATCGGAATAGCCAAAGAGACGAAATGTTGTGGAAGAGCGGAGCTGAATATCGGAGCA 122
QY      48 SerHisValIleHisLeuPheValProValSerLeuCysMetAlaLeuValValPheThr 67
DB      123 TCTCAGCTTATTCATCTATTGTCGGGTGCACATGCTGATGCTGCTGTTTACG 182
QY      68 MetAsnThrIleThrPheTyrSerGlnAsnAenGlyArgHisLeuLeuSerHisProPhe 87
DB      193 ATGACACAGATTACGTTTATAGTCAAACAAATGAGGAGCATTTACTATACATCTCTTT 242
QY      88 ValArgGluThrAspSerIleValGluLysGlyLeuMetSerLeuGlyAsnAlaLeuVal 107
DB      243 GTCCGGAAACACAGATATCTGTCAGAGGAGGATTCATGTCCTGGAATGCTCTCGTC 302
QY      108 MetLeuCysValValValLeuMetThrValLeuLeuIleValPheTyrLysTyrLysPhe 127
DB      303 ATGTGTGGCGGTGCTTCGTAAGACAGTTCTGCTGATGTTTCTTAATAATACAGTTT 362
QY      128 TyrLysLeuIleHisGlyTrpLeuIleValSerSerPheLeuLeuLeuPheLeuPheThr 147
DB      363 TATAAGCTTATTTCATGGATGGCTTATTGTCAGCAGTTTCTCTCTTTCTTCTTACT 422
QY      148 ThrIleTyrValGlnGluValLeuLysSerPheAspValSerProSerAlaLeu 165
DB      423 ACAATCTATGTCAAGAAAGTTCTGAAAGTTTCGATGTTGTCCTCCAGCGCACTA 476

```

```

RESULT 6
CD513489/c
LOCUS      884 bp      mRNA      linear      EST 06-JUN-2003
DEFINITION AGENCOURT.14365780 NIH_MGC_179 Homo sapiens cDNA clone
IMAGE:30394561 5', mRNA sequence.
ACCESSION CD513489
VERSION CD513489.1 GI:31445207
KEYWORDS EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini, Hominidae; Homo.
Mammalia; 1 to 884)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabps-remail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNI)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be

```

found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDAM473 row: e column: 02
 High quality sequence stop: 652.

FEATURES

```

Location/Qualifiers
Source      1..884
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30394561"
/tissue_type="Pituitary"
/lab_host="DH10B-ron A (T1 and T5 phage resistances)"
/notes="Organ: brain; Vector: pCMV-SPORT6.1; Site 1: EcoRV
(deprecated); Site 2: NotI; Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon cloning
). Average insert size 1.1 kb. Library was constructed by
(invitrogen). Note: this is a NIH_MGC Library."

BASE COUNT      218 a      247 c      259 g      160 t
ORIGIN
Alignment Scores:
Pred. No.:      9.09e-63      Length:      884
Score:          734.50      Matches:      164
Percent Similarity: 64.92%      Conservative: 34
Best Local Similarity: 53.77%      Mismatches: 53
Query Match:      30.89%      Indels:      54
DB:              14      Gaps:      7

US-09-043-944-1 (1-461) x CD513489 (1-884)
QY      113 ValLeuMetThrValLeu-LeuIleValPheTyrLysTyrLysPheTyrLysLeuIleHi 132
DB      785 GTGGTTATGACCATCTCTCTTGGTGGTCTCTCAAGTACCGTTGTACAGTTTCATCCA 726
QY      132 sGlyTrpLeuIleValSerSerPheLeuLeuLeuPheLeuPheThrThrIleTyrValG1 152
DB      725 TGGCTGGTTGATGATGTTCTCACTGATGCTGCTGCTTCCCTTCCCTATATATCTACCTTGG 666
QY      152 nGluValLeuLysSerPheAspValSerProSerAlaLeuLeuValLeuPheGlyLeuG1 172
DB      665 GGAAGTGCTCAAGACCTACAAATGTGCCATGGACTACCCACCCTCTTGTGACTGTCTG 606
QY      172 yAsnTyrGlyValLeuGlyMetMetCysIleHisTrpLysGlyProLeuArgLeuG1 192
DB      605 GAATCTCGGGCGAGTGGCGATGGTGTGCTACTTCCACTGGAGGGCCCTCTGTGCTGACGA 546
QY      192 nPheTyrLeuIleThrMetSerAlaLeuMetAlaLeuValPheIleLysTyrLeuProG1 212
DB      545 GGCCTACCTCATCATGATCAGTGGCTCATGGCCCTAGTGTTCATCAAGTACCTCCAGA 486
QY      212 uTrpThrValTrpPheValLeuPheValIleSerValTrpAspLeuValAlaValLeuTh 232
DB      485 GTGGTCCGCGTGGGTCTATCTCTGGCGCCATCTCTGTGTATGATCTCTGTGCTGTGCTGTG 426
QY      232 rProLysGlyProLeuArgTyrTrpLeuValGluThrAlaGlnGluArgAsnGluProIlePh 252
DB      425 TCCCAAGAGCCCTCTGAGATGCTGGTAGAAATGCCCAGAGAGAAATGAGCCCATATT 366
QY      252 eProAlaLeuIleTyrSerSerGlyValIleTyrProTyrValLeuValThr----- 289
DB      365 CCCTGCGCTGATATCTATCTCATCTGCCATGGTGTGGAGCGTTGGCATGGCAAGCTGGACCC 306
QY      270 -----AlaValGluAsnThrThrAspProArg---GluProThrSerSerAs 284
DB      305 CTCCTCTCAGGGTGGCCCTCCAGCTCCCTACGACCCCGAGATGGAGAGACCTCTCTATGA 246
QY      284 pSerAsnThrSerThrAlaPheProGlyGluAlaSerCysSerSerGluThrProLysAr 304
DB      245 CAGTTT-----GGGAG-----CCTTCATA 225
QY      304 gProLysValLysArgIleProGlnLysValGlnIleGluSerAsnThrThrAlaSerTh 324

```

```

Db      224 CCCGAAAGCTTTGAGCCTCCC-----TT 201
QY      324 rThrGlnAsnSerGlyValArgValGluArgGluLeuAlaAlaGluArgProThrValG1 344
Db      200 GACTGGCTACCCAGGGAGGAGCTGGAG----- 173
QY      344 nAspAlaAsnPheHisArgHisGluGluGluGluArgGlyValLysLeuGlyLys 364
Db      172 -----GAAGAGAGAGGAAAGGGCGCTGAAGCTTGGCTCGGGGA 135
QY      364 pPheIlePheTyrSerValLeuLeuGlyLysAlaSerSer-----TyrPheAspTrpAs 382
Db      134 CTTTCATCTTCAAGTGTGCTGGTGGGCAAGCGGCTGCCAGCGGAGGAGCTGGAA 75
QY      382 nThrThrIleAlaCysTyrValAlaIleLeuIleGlyLeuCysPheThrLeuValLeuLe 402
Db      74 TACCAGCTGGCTGCTGCTGGCCATCTCATTTGGCTTGTCTGACCCCTCCTGCTGCT 15
QY      402 uAlaValPheLys 406
Db      14 TGCTGTGTTCAG 2

RESULT 7
BX426340
LOCUS      BX426340 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
DEFINITION clone CS0G004YO20 5-PRIME, mRNA sequence.
ACCESSION BX426340
VERSION   BX426340.1 GI:30772565
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLES    Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
JOURNAL   Full-length cDNA libraries and normalization
COMMENT   Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 595.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0AG004BH10QP1&cluster=595.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0AG004BH10QP1.
FEATURES
Location/Qualifiers
1..964
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DG004YO20"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 195 a 266 c 271 g 230 t
ORIGIN
Alignment Scores:
Pred. No.: 1,16e-62 Length: 964
Score: 734.00 Matches: 162
Percent Similarity: 64.72% Conservative: 49
Best Local Similarity: 49.69% Mismatches: 69
Query Match: 30.87% Indels: 47

```

RESULT 8

```

DB:      13          Gaps:      5
US-09-043-944-1 (1-461) x BX426340 (1-964)
QY      75 SerGlnAsnAsnGlyArgHisLeuLeuSerHisProPheValArgGluThrAspSerIle 94
Db      105 TCCAGCAGCTTTGGGAGAGCTGAGCATCTACAGCCATTCCATGAGGACACACCCCTCGGTG 164
QY      95 ValGluLysGlyLeuMetSerLeuGlyAsnAlaLeuValMetLeuMetLeuCysValValVal 114
Db      165 GGCAGCGCCTCTCAACTCCGTGCTGAACACCTCATCATCATCATCATCATCATCATCAT 224
QY      115 MetThrValLeuLeuValPheTyrLysTyrLysPheTyrLysLeuLeuHisGlyTyr 134
Db      225 ATGACATCTTCTTGGTGTCTCTACAAGTACCGCTGCTCAAGTTCATCAAGTTCATCA 284
QY      135 LeuIleValSerSerPheLeuLeuLeuPheLeuPheThrThrIleTyrValGlnGluVal 154
Db      285 TTGATCATGTCTTCACTGATGCTGCTTCCTTCCCTTACCTATATCTTACCTTGGGGA 344
QY      155 LeuLysSerPheAspValSerProSerAlaLeuLeuValLeuPheGlyLeuGlyAsnTyr 174
Db      345 CTCAGACCTACATGTGCCCATGAGCTACCCACCCCTCTTCTGACTGCTCTGGAACCTC 404
QY      175 GlyValLeuGlyMetMetCysIleHisTyrLysGlyProLeuArgLeuGlnGlnPheTyr 194
Db      405 GGGGAGTGGGCGATGCTGTCATCCATGGAAGGGCCCTCTGCTGCTGAGCAGGCGCTAC 464
QY      195 LeuIleThrMetSerAlaLeuMetAlaLeuValPheIleLysTyrLeuProGluThrThr 214
Db      465 CTCATCATGATCAGTGGCTCATGCCCTAGTGTTCATCACTATCTTCCAGTACCTCC 524
QY      215 ValTrpPheValLeuPheValIleSerValTrpAspLeuValAlaValLeuThrProLys 234
Db      525 CGCTGGGTCTCTCTGGGCGCCATCTCTGTGTATGATCTCTGCTGCTGCTTCTGTCCCAA 584
QY      235 GlyProLeuArgTyrLeuValGluThrAlaGlnGluArgAsnGluProIlePheProAla 254
Db      585 GGGCCTCTGAGAATGCTGTGAGAACTGCCAGGAGAGAAATGAGCCCATATTTCCTGTC 644
QY      255 LeuIleTyrSerSerGlyValIleTyrProTyrValLeuValThrAlaValGluAsnThr 274
Db      645 CTGATATACTCATCTGCCATGCTGGTGTGGACGGTGGCATGGCGGAATGNN----- 692
QY      275 ThrAspProArgGluProThrSerSerAspSerAsnThrSerThrAlaPheProGlyGlu 294
Db      693 -----GACCCCTCTCTCAGGGT----- 710
QY      295 AlaSerCysSerSerGluThrProLysArgProLysVal-LysArgIleProGlnLysVa 314
Db      711 -----GCCCTCCAGCTCCCTTACGCCGAGATGGAAGAAGACTCTCTATGACAGT 761
QY      314 IglIleGluSerAsnThrThrAlaSerThrThr-----GlnAsnSerGlyVa 330
Db      762 TTTGGGAGCCTTCAATACCCGAGCTTTTGAGCCTCCCTTGAAGCTACCTACCCAGGGGA 821
QY      330 IargValGluArgGluLeuAlaAlaGluArgProThrValGlnAspAlaAsnPheHisAr 350
Db      822 GGAGCTGGAG----- 831
QY      350 GHisGluGluGluArgGlyValLysLeuGlyLeuGlyAspPheIlePheTyrSerVa 370
Db      832 ----GAAGAGAGGAAAGGGCGTGAAGCTTGGCTCGGGGACTTCTCTTACAGTGT 887
QY      370 IleuLeuGlyLysAlaSer-----SerTyrPheAspTrpAsnThrThrIleAlaCysTy 388
Db      888 GCTGTGGGCAAGCGGCTGCACGGGCACGGGG-GACTGGGAATACACGCTCGCTGCTT 946
QY      388 rValAlaIleLeuIle 393
Db      947 CGTGGCCATCTTCATT 962

```


Db 503 TACAAGTATCGATCTACAAAGTTTCATCCATGGCTGGCTGATCATGCTCTCCCTGATGCTC 562
 Qy 143 LeuPheLeuPheThrThrLeuValGlnGluValLeuLysSerPheAspValSerPro 162
 Db 563 CTCTCTTGTTCACCTACATCTACCTCGGGGAAGTGTCAAGACCTACATATGGCCATG 622
 Qy 163 SerAlaLeuLeuValLeuPheGlyLeuGlyAsnTyrGlyValLeuGlyMetMetCysIle 182
 Db 623 GACTATCCACACATCTCTCTGCTGCTGGAAGTGTGGGCGAGTGGGCGATGGTGCATC 682
 Qy 183 HisTrpLysGlyProLeuArgLeuGlnGlnPheTyrLeuLeuLeuMetSerAlaLeuMet 202
 Db 683 CACTGGAAGGGCGCTCTGGTGTCTGACAGAGCTTACCTTATTGTGATCAGCGCACTCATG 742
 Qy 203 AlaLeuValPheLeuLeuValLeuPheGlyValLeuPheValLeuValLeuValLeu 222
 Db 743 GCCCTGTGTTCATCAAGTACCTGCCGAGTGTCTGCTGGTGCATCTTGGTGGTGCATC 802
 Qy 223 SerValTrpAspLeuValAlaValLeuThrProLysGlyProLeuArgTyrLeuValGlu 242
 Db 803 TCTGTGTACGATCTCTGGTGGCTGTGTGCCCAAGGGCCACTGAGGATGCTGGTGGAA 862
 Qy 243 ThrLagGlnGluArgAsnGluProLeuPheProLeuLeuLeuLeuLeuLeuLeu 258
 Db 863 ACTGCCAGGAGAGAAATGAGCCCATATTTCTGCCCTGATATCTCA 910
 RESULT 10
 A1675803 661 bp mRNA linear EST 17-DEC-1999
 LOCUS W997a12.x1 NCI CGAP Pr28 Homo sapiens cDNA Clone IMAGE:2313598 3
 DEFINITION similar to SW:PSN2_HUMAN P49810 PRESENILIN 2 ; mRNA sequence.
 ACCESSION A1675803
 VERSION A1675803.1 GI:4876283
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 661)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
www-bio.lnl.gov/bbrp/image/image.html
 Insert Length: 1317 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 461.
 Location/Qualifiers
 1..661
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2313598"
 /sex="male"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="NCI-CGAP_Pr28"
 /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
 with a modified polylinker; Plasmid DNA from the
 normalized library NCI CGAP Pr22 was prepared, and ss
 circles were made in vitro. Following HAP purification,
 this DNA was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from a pool
 of 5,000 clones made from the same library (cloneIDs

985608-986759, 1101192-1101959, and 1217928-1220615).
 Subtraction by Bento Soares and M. Fatima Bonaudo. "
 BASE COUNT 132 a 190 c 166 g 171 t 2 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 9,48e-60 Length: 661
 Score: 702.50 Matches: 135
 Percent Similarity: 79.26% Conservative: 37
 Best Local Similarity: 62.21% Mismatches: 44
 Query Match: 29.54% Indels: 1
 DB: 9 Gaps: 1
 US-09-043-944-1 (1-461) x A1675803 (1-661)
 Qy 43 LeuLysTyrGlyAlaSerHisValIleHisLeuPheValProValSerLeuCysMetAla 62
 Db 8 CTCAAATACGGAGCAAGCAGTCATGCTGTGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 67
 Qy 63 LeuValValPheThrMetAsnThrIleThrPheTyrSerGlnAsnAsnGlyArgHisLeu 82
 Db 68 GTGGTGGTAGCCACCATCAAGTCTGTGCGCTTCTACAGAGAGAAATGGACAG--CTC 124
 Qy 83 LeuSerHisPropheValArgGluThrAspSerIleValGluLysGlyLeuMetSerLeu 102
 Db 125 ATCTACAGCCATTCACTGAGGACACACCTCGGTGGGCCAGCGCTCTCTCAACTCCGCTG 184
 Qy 103 GlyAsnAlaLeuValMetLeuCysValValValLeuMetThrValLeuLeuValPhe 122
 Db 185 CTGAACACCTCATCATGATCAGCGTCATCGGTGTATGACCATCTTCTGGTGTGCTC 244
 Qy 123 TyrLysTyrLysPheTyrLysLeuLeuHisGlyTyrLeuLeuValSerSerPheLeuLeu 142
 Db 245 TACNAGTACCGCTGTGTCAAGTTCATCCATGCTGCTGTGATCATGCTCTTCACTGATGCTG 304
 Qy 143 LeuPheLeuPheThrThrIleTyrValGlnGluValLeuLysSerPheAspValSerPro 162
 Db 305 CHTGTCTCTTCCACCTATATCTACCTGTTGGGGAAGTGTCTCAAGCTCATATGTGCCCATG 364
 Qy 163 SerAlaLeuLeuValLeuPheGlyLeuGlyAsnTyrGlyValLeuGlyMetMetCysIle 182
 Db 365 GACTACCCACCTCTTGTGCTGACTGTCTGGAAGTCTGGGGGAGTGGGAGTGTGTGCTATC 424
 Qy 183 HisTrpLysGlyProLeuArgLeuGlnPheTyrLeuLeuThrMetSerAlaLeuMet 202
 Db 425 CACTGGAAGGGCGCTCTGCTGTGTCAGAGCGCTACCTCATCATCATCATCATCATCATCATG 484
 Qy 203 AlaLeuValPheIleLysTyrLeuProGluTyrThrValTrpPheValLeuValIle 222
 Db 485 GCCTAGTGTTCATCAAGTACCTCCAGAGTGTGTCGGCTGGGTCTATCTCTGGGGCCCATC 544
 Qy 223 SerValTrpAspLeuValAlaValLeuThrProLysGlyProLeuArgTyrLeuValGlu 242
 Db 545 TCTGTGTATGATCTCGTGGCTGTGTGTGTCCTCCNAGGGCGCTCTGAGAATGTGTGTAGAA 604
 Qy 243 ThrAlaGlnGluArgAsnGluProIlePheProAlaLeuLeuLeuTyrSerSer 259
 Db 605 ACTGCCAGGAGAGAAATGAGCCCATATTTCTGCCCTGATATCTATCTATCT 655
 RESULT 11
 BU614355 697 bp mRNA linear EST 20-FEB-2003
 LOCUS UI-M-EV0-cbf-1-22-0-UI.r1 NIH EMAP_EV0 Mus musculus cDNA clone
 DEFINITION UI-M-EV0-cbf-1-22-0-UI 5', mRNA sequence.
 ACCESSION BU614355
 VERSION BU614355.1 GI:23280570
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 697)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

```

TITLE
JOURNAL
COMMENT
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1. .697
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="UI-M-EV0-cbf-1-22-0-UI"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP EV0"
/note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is GTGGCTGGAA. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP): 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT 160 a 173 c 158 g 204 t
ORIGIN

Alignment Scores:
Pred. No.: 2-29e-59 Length: 697
Score: 699.00 Matches: 135
Percent Similarity: 77.83% Conservative: 44
Best Local Similarity: 58.70% Mismatches: 47
Query Match: 29.39% Indels: 4
DB: 13 Gaps: 2

US-09-043-944-1 (1-461) x BU614355 (1-697)

QY 43 LeuLysTyGlyAlaSerHisValIleHisLeuPheValProValSerLeuCysMetAla 62
Db 7 TTGAATATGGAGCAAGCATGTTCATCATGCTCTTTGTCCTCCGTCGACCTCTGCATGGTC 66
QY 63 LeuValValPheThrMetAsnThrIleThrPheTyrSerGlnAsnAnglyArgHisLeu 82
Db 67 GTGCTGTGGCCACCAATCAATCAGTCAGCTTCTATACCGGAGGACGCTCAG---CTA 123
QY 83 LeuSerHisProPheValArgGluThrAspSerIleValGluLysGlyLeuMetSerLeu 102
Db 124 ATCTACACCCCTTACAGNAGACACTGAGACTGTAGGCCAAGAGCCCTGCACTGCATC 183
QY 103 GlyAsnAlaLeuValMetLeuCysValValValLeuMetThrValLeuLeuIleValPhe 122
Db 184 CTGAATGCGGCCCATCATGATCAGTGTTCATTGTATTATGACCATCTCTCTGGTGGTCTG 243
QY 123 TTTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTT 142
Db 244 TATAAATACAGGTGTCACAGGTGTCACAGGTGTCACAGGTGTCACAGGTGTCACAGGTGTC 303
QY 143 LeuPheLeuPheThrThrIleTyrValGlnGluValLeuLysSerPheAspValSerPro 162

```

```

Db 304 CTGTTCTTTTTCGTTTCATTACTTAGGGGAAGTATTTAAGACCTACAAATGTCGCGTG 363
QY 163 SerAlaLeuValLeuPheGlyLeuGlyAsnTyrGlyValLeuGlyMetMetCysIle 182
Db 364 GACTAGCTACAGTAGCAGTCCCTAAATCTGGAATTTGGTGTGGTGGGATGATGCCATC 423
QY 183 HisTrpGlyGlyProLeuArgLeuGlnPheTyrIleuIleThrMetSerAlaLeuMet 202
Db 424 CACTGAAAGGCCCCCTTCGACTGCAGCAGCGCTATCTCATTATGATCAGTGCCTCATG 483
QY 203 AlaLeuValPheIleLysTyrLeuProGluTyrThrValTyrPheValLeuPheValIle 222
Db 484 GCCTGTGATTTATCAAGTACCTCCCGAATGACCGCATGGCTCATCTTTGGTGTGATT 543
QY 223 SerValTrpAspLeuValAlaValLeuThrProLysGlyProLeuArgTyrLeuValGlu 242
Db 544 TCAGTATATGATTTGGTGGCTGTTTATGTCNCAAGGCCCATCTTGGTATGCTGTTGAA 603
QY 243 ThrAlaGlnGluArgAsnGluProIlePheProAlaLeuIleTyrSerGlyValIle 262
Db 604 ACAGCTCAGGAAAGAAATGAGACTCTCTTCCAGCTCTTATCTATCTCTCAACANTGGTG 663
QY 263 TyrProTyrValLeuValThrAlaValGlu 272
Db 664 TGG-----TTGGTGAATATGGCTGAA 684

RESULT 12
AI925372
LOCUS 703 bp mRNA linear EST 02-SEP-1999
DEFINITION wn53d06.x1 NCT_CGAP Lu19 Homo sapiens cDNA clone IMAGE:249163 3'
similar to SW:FSN2_HUMAN P49810 PRESENILIN 2 ; mRNA sequence.
ACCESSION AI925372
VERSION AI925372.1 GI:5661336
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 703)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 459.
Location/Qualifiers
1. .703
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:249163"
/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NCI_CGAP_Lu19"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
oiled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and

```

cloned into the Not I and Eco RI sites of the modified pTV3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 138 a 203 c 177 g 184 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 2.6e-59 Length: 703
Score: 698.50 Matches: 134
Percent Similarity: 78.73% Conservative: 40
Best Local Similarity: 60.63% Mismatches: 46
Query Match: 29.37% Indels: 1
DB: 9 Gaps: 1

US-09-043-944-1 (1-461) x A1925372 (1-703)

QY 43 LeuLysTyrGlyAlaSerHisValIleHisLeuPheValProValSerLeuCysMetAla 62
Db 8 CTCAATACCTAGGAGACGATGATCATGCTGTTGTGCTGTCTCTGTGATGATC 67
QY 63 LeuValValPheThrMetAsnThrIleThrPheTyrSerGlnAsnGlyArgHisLeu 82
Db 68 GTGGTGTAGCCACCATCAAGTCTGTCGGCTTCTACACAGAGAAGATGGACAG--CTC 124
QY 83 LeuSerHisProPheValArgGluThrAspSerIleValGluLysGlyLeuMetSerLeu 102
Db 125 ATCTACAGCGCATTCACCTAGGACACACCCCTCGGTGGCCAGCGCCCTCACTCGGTG 184
QY 103 GlyAsnAlaLeuValMetLeuCysValValValLeuMetThrValLeuLeuValPhe 122
Db 185 CTGAACACCCCTCAATCATGATGACGGTCTGCTGTTGATGACCATCTCTTCTGTTGGTCTC 244
QY 123 TyrLysTyrLysPheTyrLysLeuIleHisGlyTyrLeuIleValSerSerPheLeuLeu 142
Db 245 TACAAGTACCGTGTCTACAAGTTCTCATCGGCTGGTTGATCATGCTTCTCACTGATGCTG 304
QY 143 LeuPheLeuPheThrThrIleTyrValGlnGluValLeuLysSerPheAspValSerPro 162
Db 305 CTGTGTCCTCTTCACTATATCTACTTGGGGAAGTGTCTCAAGACCTACATGTGGCCATG 364
QY 163 SerAlaLeuLeuValLeuPheGlyLeuGlyAsnTyrGlyValLeuGlyMetCysIle 182
Db 365 GACTACCCACCCCTCTTGTGCTGACTGTCTGGACTTGGGGGAGTGGGCGATGGTGTGATC 424
QY 183 HisTyrLysGlyProLeuArgLeuGlnGlnPheTyrLeuIleThrMetSerAlaLeuMet 202
Db 425 CACTGGAAGGGCCCTCTGTGGTGTCTGACAGCGCTTACCTCATCATGATCATGTGGCTCATG 484
QY 203 AlaLeuValPheIleLysTyrLeuProGluTyrThrValTyrPheValLeuPheValIle 222
Db 485 GCCTAGTGTTCATCAAGTACCTCCAGAGTGGTCCCGGTGGGTCATCTCTGNGCGCCATC 544
QY 223 SerValTyrAspLeuValAlaValLeuThrProLysGlyProLeuArgTyrLeuValGlu 242
Db 545 TCTGTGTATGATCTCTGGGTGTCTGTGTCCTCCAAAGGGCTCTGAGAATGCTGTGAGAA 604
QY 243 ThrAlaGlnGluArgAsnGluProIlePheProAlaLeuIleTyrSerSerGlyValIle 262
Db 605 ACTGCCAGGAGATATGAGCCCATATTCCTGCCCTGATATATCTCATCTGCGCCATGCTG 664
QY 263 Tyr 263
Db 665 TGG 667

RESULT 13

BX426339/c

LOCUS BX426339 974 bp mRNA linear EST 15-MAY-2003
DEFINITION BX426339 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
clone CSODG004YO20 3-PRIME, mRNA sequence.

ACCESSION BX426339

VERSION BX426339.1 GI:30770561

KEYWORDS EST.

SOURCE

ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 974)

AUTHORS

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

JOURNAL

Unpublished

COMMENT

Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster S95.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cqi-bin/cluster.cgi?seq=CS0AG004BH1ONP1&cluster=595.f. Contact :
Peng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0AG004BH1ONP1.

FEATURES

source

1..974
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D004YO20"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 247 a 263 c 275 g 185 t 4 others
ORIGIN

Alignment Scores:

Pred. No.: 4.15e-58 Length: 974
Score: 688.50 Matches: 163
Percent Similarity: 61.94% Conservative: 29
Best Local Similarity: 52.58% Mismatches: 57
Query Match: 28.95% Indels: 62
DB: 13 Gaps: 8

US-09-043-944-1 (1-461) x BX426339 (1-974)

QY 129 LysLeuIleHisGlyTyrTrpLeuIleValSerSerPheLeuLeuLeuPheLeuPheThrThr 148
Db 963 AAGTTTCATCGCTGGCTGGTGGTATCATCTTAACGTGAGCTGTGTTCCCTTCCCTTAT 904
QY 149 IleTyrValGlnGluValLeuLysSerPheAspValSerProSerAlaLeuValLeu 168
Db 903 ATCTACCTTGGGGAAGTGTCTCAAGACCTACATGTGGCCATGACTACCCACCTCTTG 844
QY 169 PheGlyLeuGlyAsnTyrGlyValLeuGlyMetMetCysIleHisTyrLysGlyProLeu 188
Db 843 CTAACTGTCTGGAACCTTCGGGGCAGTGGGATGTTGTGTCATCCACTGGAAGGGCCCTG 784
QY 189 ArgLeuGlnGlnPheTyrLeuIleThrMetSerAlaLeuMetAlaLeuValPheIleLys 208
Db 783 GTGCTGAGCAGGCGCTTACCTCATCATGATCAGTGGCTCATGGCCCTAGTGTTCATCAAG 724
QY 209 TyrLeuProGluTyrThrValTyrPheValLeuPheValIleSerValTyrAspLeuVal 228
Db 723 TACCTCCAGAGTGTGTCGGTGGTATCTCTGGGGCCCATCTCTGTGTATGATCTCGTG 664
QY 229 AlaValLeuThrProLysGlyProLeuArgTyrLeuValGluThrAlaGlnGluArgAsn 248
Db 663 GCTGTGCTGTGTCCTCCAAAGGGCCCTCTGAGAATCTGTGATAGAACTGCCAGGAGAAAT 604
QY 249 GluProIlePheProAlaLeuIleTyrSerSerGlyValIleTyrProTyrValLeuVal 268
Db 249 GluProIlePheProAlaLeuIleTyrSerSerGlyValIleTyrProTyrValLeuVal 268

```

Db      603 GAGCCCATATTCCTCCCTCGATATATATCTATCTATCCCATGCTGTGGACGGTGGCATGGCG 544
QY      269 Thr-----AlaValGluAsnThrThrAspProArg-----GluPro 280
Db      543 AAGCTGAGACCCCTCTCTCAGGTGCTCCCTCCAGCTCCCTACACCCGGAGATGGAAGAA 484
QY      281 ThrSerSerAspSerAsnThrSerThrAlaPheProGlyGluAlaSerCysSerSerGlu 300
Db      483 GACTCCTCATGACATTT-----GGGAG-----460
QY      301 ThrProLysArgProLysValLysArgIleProGlnLysValGlnIleGluSerAsnThr 320
Db      459 ---CCTTCATACCCCGAAGCTTTGAGCCCTCC-----430
QY      321 ThrAlaSerThrThrGlnAsnSerGlyValArgValGluArgGluLeuAlaAlaGluArg 340
Db      429 -----TTGACTGCTACCCAGGAGGAGCTGGAG-----400
QY      341 ProThrValGlnAspAlaAsnPheHisArgHisGluGluGluArgGlyValLysLeu 360
Db      399 -----GAGAGGAGGAAGAGGGCGTGAAGCTT 373
QY      361 GlyLeuGlyAspPheIlePheTyrSerValLeuLeuGlyLysAla-----SerSerTyr 378
Db      372 GGCCTCGGGGACTTCTCTCAGTGTGCTGTGGCGAAGCGCGTCCACGGCGACGG 313
QY      379 PheAspTrpAsnThrThrIleAlaCysTyrValAlaIleLeuIleGlyLeuCysPheThr 398
Db      312 GG-GACTGGATACACAGCTGGCTGCTCTGTGGCCATCCTCATTTGCTGTGACC 254
QY      399 LeuValLeuLeuAlaValPheLysArgAlaLeuProAlaLeuGln-----413
Db      253 CTCCTGCTGTGCTGTGTTCAAGAAGCGCTGCGCGCCCTCCCATCTCCATCACGTTT 194
QY      414 -----PheProPheSerProAspSer 420
Db      193 GGGCTCATCTTTTACTTCTCCACGGACAAAC 164

RESULT 14
LOCUS   A1831581
DEFINITION 680 bp mRNA linear EST 21-DEC-1999
similar to SW:PSN2_HUMAN P49810 PRESENILIN 2, mRNA sequence.
ACCESSION A1831581
VERSION   A1831581.1 GI:5452252
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 680)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Prepared by: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1711 sfd Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 456.
Location/Qualifiers
1. .680
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

FEATURES
source

```

/clone="IMAGE:2405191"
/tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"
/dev stage="adult"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NCI CGAP Lu19"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 137 a 191 c 173 g 178 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 1.24e-57 Length: 680
Score: 681.50 Matches: 131
Percent Similarity: 78.28% Conservative: 42
Best Local Similarity: 59.28% Mismatches: 47
Query Match: 28.66% Indels: 1
DB: Gaps: 1

US-09-043-944-1 (1-461) x A1831581 (1-680)

QY 43 LeuLysTyrGlyAlaSerHisValIleHisLeuPheValProValSerLeuCysMetAla 62
Db 9 CTCAAAATCTATGATCGAAGCAGCTGATCATGCTGTTGTGCTCTCTCTCTGTCATGTC 68
QY 63 LeuValValPheThrMetAsnThrIleThrPheTyrSerGlnAsnAsnGlyArgHisLeu 82
Db 69 GTGGTGTAGCCACCATCAAGTCTGTGGCTTCTACACAGAGAAGATGGACAG---CTC 125
QY 83 LeuSerHisProPheValArgGluThrAspSerIleValGluLysGlyLeuMetSerLeu 102
Db 126 ATCTACACGGCCATTCTACTGAGGACACACCCCTCGGTGGCCAGCGCCTCTCACTCCGTG 185
QY 103 GlyAsnAlaLeuValMetLeuCysValValValLeuMetThrValLeuLeuIleValPhe 122
Db 186 CTGAACACCTTCATCATGATCATGCTCATCTGTTGTTATGACCATCTTCTTGTGTGCTC 245
QY 123 TyrLysTyrLysPheTyrLysLeuIleHisGlyTrpLeuIleValSerSerPheLeuLeu 142
Db 246 TACAAGTACCCTGCTACAGTTTCATCATGGCTGGTTGATCATGCTCTTCACTGATGCTG 305
QY 143 LeuPheLeuPheThrThrIleTyrValGlnGluValLeuLysSerPheAspValSerPro 162
Db 306 CTGTTCTCTTTCACCTATATCTACTTGGGAAGTGTCTAAGACCTACAATGTGGCCATG 365
QY 163 SerAlaLeuLeuValLeuPheGlyLeuGlyAsnTyrGlyValLeuGlyMetMetCysIle 182
Db 366 GACTACCCACCCCTCTTCTGACTGCTGGAACCTTCGGGGCAGTGGGATGTTGTGTCATC 425
QY 183 HisTrpLysGlyProLeuArgLeuGlnPheTyrLeuIleThrMetSerAlaLeuMet 202
Db 426 CACTGGAAGGGCCCTCTCTGCTGCTGAGGAGTGGCGGTGGTGGTGGTGGTGGTGGTGGT 485
QY 203 AlaLeuValPheIleLysTyrLeuProGluThrPheValThrPheValLeuPheValIle 222
Db 486 GCCCTAGTGTTCATCAAGTACTCTCCAGTGGTGGCGGTGGTGGTGGTGGTGGTGGTGGT 545
QY 223 SerValTrpAspLeuValAlaValLeuThrProLysGlyProLeuArgTyrLeuValGlu 242
Db 546 TCTGTGTATGATCTCTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 605
QY 243 ThrAlaGlnGluArgAsnGluProIlePheProAlaLeuIleTyrSerSerGlyValIle 262
Db 606 ACTGCCAGGAGAGATGATGAGCCCATATTCCCTGCTGATATATCTCATCTGACATGTTG 665

Search completed: December 11, 2003, 08:21:17
Job time : 2296 secs

Alignment Scores:			
Pred. No.:	2.02e-56	Length:	881
Score:	671.00	Matches:	145
Percent Similarity:	63.01%	Conservative:	39
Best Local Similarity:	49.66%	Mismatches:	64
Query Match:	28.22%	Indels:	44
DB:	14	Gaps:	5

US-09-043-944-1 (1-461) x CD358381 (1-881)			
Qy	86	ProPheValArgGluThrAspSerIleValGluLysGlyLeuMetSerLeuGlyAsnAla	105
Db	2	CCATTCTCAGGACACACCTCGGTGGCGCAGGGCTCTCAACTCCGTGTGACACC	61
Qy	106	LeuValMetLeuCysValValLeuMetThrValLeuLeuIleValPheTyrIysTyr	125
Db	62	CTCATCATGATCAGCGTCATCGTGGTTATGACCATCTTCTTGTTGGTCTCTACAGTAC	121
Qy	126	LysPheTyrIysLeuIleHisGlyTyrLeuIleValSerPheLeuLeuLeuPheLeu	145
Db	122	CGCTGTACAAAGTTCACCAAGGCTGGTTGATCATGTCITCACTGAGCTGCTGTTCCTC	181
Qy	146	PheThrThrIleTyrValGlnGluValLeuLysSerPheAspValSerProSerAlaLeu	165

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 11, 2003, 06:24:06 ; Search time 339 Seconds
(without alignments)
3670.917 Million cell updates/sec

Title: US-09-043-944-1

Perfect score: 2378

Sequence: 1 MFSTRQGGGADAEHTV.....YYINSLFLPFLCINFIIS 461

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349713017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/usgpt0 spool /US09043944/runat.11122003.062359.3663/app.query.fasta_1.647
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFPLX=ring -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DLOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09043944 @CGN 1.1.490 @runat.11122003.062359.3663 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2378	100.0	1500	18	AAT60306
2	2052.5	86.3	4137	24	AAL47322
3	994.5	41.8	2765	18	AAT85332
4	993.5	41.8	1488	18	AAT87402
5	993.5	41.8	1703	19	AAV17357
6	993.5	41.8	2764	19	AAV17358
7	993.5	41.8	2764	24	AAL47323
8	993.5	41.8	2765	17	AAT40028
9	993.5	41.8	2765	19	AAV04666
10	993.5	41.8	2765	22	AAD18120
11	993.5	41.8	2765	22	AAT74993
12	991	41.7	1750	18	AAT59536
13	991	41.7	1762	18	AAT59535
14	989.5	41.6	3056	24	ABK83912
15	989.5	41.6	3086	17	AAT40029
16	989.5	41.6	3086	19	AAV04667
17	989	41.6	1964	17	AAT40030
18	989	41.6	1964	19	AAV04668
19	989	41.6	2691	18	AAT64819
20	988.5	41.6	1895	17	AAT40043
21	988.5	41.6	1911	18	AAT63207
22	988.5	41.6	1914	18	AAT75576
23	988.5	41.6	2048	23	ABL29237
24	986.5	41.5	2764	19	AAV29525
25	985.5	41.4	1404	20	AAV03246
26	983	41.3	1750	19	AAV03246
27	979.5	41.2	2765	18	AAT85333
28	979	41.2	1392	19	AAV75761
29	979	41.2	1404	20	AAV90184
30	979	41.2	1404	22	AAD10303
31	970.5	40.8	2276	18	AAT87426
32	969	40.7	1404	24	AAD27443
33	967	40.7	2229	17	AAT40031
34	967	40.7	2229	19	AAV04669
35	967	40.7	2236	18	AAT51253
36	964.5	40.6	1705	25	ACA56467
37	959	40.3	2144	21	AAZ40668
38	959	40.3	2144	25	ACC44830
39	959	40.3	2236	19	AAV75762
40	959	40.3	2236	22	AAD18121
41	959	40.3	2236	22	AAT74994
42	958.5	40.3	2527	22	AAH98480
43	950.5	40.0	1983	21	AAZ40670
44	949	39.9	1347	24	AAD27446
45	947	39.8	1347	24	AAD27447

ALIGNMENTS

RESULT 1
AAT60306
ID AAT60306 standard; DNA; 1500 BP.
XX AAT60306;
AC AAT60306;
XX
XX 11-JUL-1997 (first entry)
DT
XX
XX Caenorhabditis elegans SEL-12 gene.
DE
XX
XX SEL-12; presenilin; transgenic animal; Alzheimer's disease;
KW animal model; therapy; diagnosis; ss.
XX
XX Caenorhabditis elegans.
XX
XX
XX Key Location/Qualifiers
FH CDS 22..1407
FT


```

FT      /*tag= a
XX      WO9711956-A1.
XX      03-APR-1997.
XX      27-SEP-1996; 96WO-US15727.
XX      27-SEP-1995; 95US-0004387.
XX      (UYCO ) UNIV COLUMBIA NEW YORK.
XX      Greenwald I, Levitan D;
XX      WPI; 1997-212841/19.
XX      P-PSDB; AAW14006.
XX
XX      New isolated SEL-12 presenilin gene - used to develop products for
XX      the diagnosis and prophylactic or therapeutic treatment of disorders
XX      such as Alzheimer's disease
XX
XX      Claim 11; Page 75-76; 99pp; English.
XX
XX      The sel-12 gene (AAW60306) of Caenorhabditis elegans (CE) encodes a
XX      polypeptide (AAW14006) that displays about 50% amino acid sequence
XX      identity to human presenilin PS1 and PS2 proteins associated with
XX      Alzheimer's disease (AD). Like presenilins, sel-12 is widely
XX      expressed in neural and non-neural cells. The gene was identified
XX      by screening for suppressors of the 'Multivulva' phenotype of CE
XX      caused by an allele of lin-12 that causes constitutive LIN-12
XX      activation. 2 Recessive suppressors, arl31 and arl33, proved to be
XX      alleles of the new gene (sel) means suppressor and/or enhancer of
XX      lin-12). The gene can be used to produce wild-type or mutated (see
XX      also AAW14007) SEL-12 in host cells, in methods for screening for
XX      extragenic suppressors or enhancers of a SEL-12 allele, and in the
XX      development of transgenic animals, esp. transgenic CE, useful in
XX      identifying cpds. that ameliorate AD.
XX
XX      SQ Sequence 1500 BP; 403 A; 312 C; 314 G; 471 T; 0 other;
XX
XX      Alignment Scores:
XX      Pred. No.: 3,946-230 Length: 1500
XX      Score: 2378.00 Matches: 461
XX      Percent Similarity: 100.00% Conservative: 0
XX      Best Local Similarity: 100.00% Mismatches: 0
XX      Query Match: 100.00% Indels: 0
XX      DB: 18 Gaps: 0
XX
XX      US-09-043-944-1 (1-461) x AAT60306 (1-1500)
XX
XX      1 MetProSerThrArgGlnGlnGluGlyGlyAlaAspAlaGluThrHisThrVal 20
XX      22 ATGCTTCCACAGAGAGACACAGAGGGGGAGGTGCAGATGCGGAACACATACCGTT 81
XX
XX      21 TyrGlyThrAsnLeuIleThrAsnArgAsnSerGlnGluAspGluAsnValGluGlu 40
XX      82 TACGGTACAAATCTCATACAAATCGAATAGCCAAAGAGAGAGAAATGTTGTGGAAGAA 141
XX
XX      41 AlaGluLeuLysTyrGlyAlaSerHisValIleHisIlePheValProValSerLeuCys 60
XX      142 GCGGAGCTGGAATACGAGACATCTCACGTTATTCATCTATTGTGCGCGGTGCTACTATGC 201
XX
XX      61 MetAlaLeuValPheThrMetAsnThrIleThrPheTyrSerGlnAsnAsnGlyArg 80
XX      202 ATGGCTCTGTTGTTTACGATGACACGATTTACGTTTATAGTCAAAACAATGGAAGG 261
XX
XX      81 HisLeuLeuSerHisProPheValArgGlnThrAspSerIleValGluLysGlyLeuMet 100
XX      262 CATTTACTATCATCCCTTTTGTCCGGAACACAGATATCGTTGGAAGGGGATGATG 321
XX
XX      101 SerLeuGlyAsnAlaLeuValMetLeuValMetLeuValValMetThrValLeuLeu 120
XX      322 TCACITGGGAATGCTCTGTCATGTTGTGCGGTGCTGCTTCTGATGACAGTCTGCTGATT 381

```

```

QY      121 ValPheTyrLysTyrLysPheTyrLysLeuIleHisGlyTyrLeuIleValSerSerPhe 140
DB      382 GTTTCATATAAATACAAAGTTTTATTAAGCTTATTCATGAAGATGCTTATTTGTCAGCAGTTTT 441
QY      141 LeuLeuLeuPheLeuPheThrThrIleTyrValGlnGluValLeuLysSerPheAspVal 160
DB      442 CTTCTCTTTTCTTATTCACATCAATCTATGTGCAAGAAAGTTCTGAAAGTTTCGATGTG 501
QY      161 SerProSerAlaLeuLeuValLeuPheGlyLeuGlyAsnTyrGlyValLeuGlyMetMet 180
DB      502 TCTCCAGCGCTATTTGTTTGGACTGGGTAACTATGGAGTTCTCGGAATGAG 561
QY      181 CysIleHisTyrLysGlyProLeuArgLeuGlnGlnPheTyrLeuIleThrMetSerAla 200
DB      562 TGTATACATGGAAAGGTCCATTGCGTCTGCACAGTTCTACCTTATTACAATGTCTCA 621
QY      201 LeuMetAlaLeuValPheIleLysTyrLeuProGluTyrThrValTyrPheValLeuPhe 220
DB      622 CTAAATGGCTCTGCTCTTATCAAGTACCTACAGAAATGAGCTGTGTGTTTGTGCTGTT 681
QY      221 ValIleSerValTyrAspLeuValAlaValLeuThrProLysGlyProLeuArgTyrLeu 240
DB      682 GTTATCTCGTTTGGATCTGTTTCCGCTGCTCACACCAAAAGGACCATTTGAGATATTG 741
QY      241 ValGluThrAlaGlnGluArgAsnGluProIlePheProAlaLeuIleTyrSerSerGly 260
DB      742 GTGGAATCTGCACAGAGAGAAACGAGCCAAATTTCCCGCGCTGATTTATTTCGTCTGGA 801
QY      261 ValIleTyrProTyrValLeuValThrAlaValGluAsnThrThrAspProArgGluPro 280
DB      802 GTCATCTATCTCTGCTGCTTCTGCTGAGTTGAAACACACAGACAGACCCCGTGAACCG 861
QY      281 ThrSerSerAspSerAsnThrSerThrAlaPheProGlyGluAlaSerCysSerSerGlu 300
DB      862 AGTCGTGAGACTCAAAATCTTCTACAGCTTTCTCTGGAGAGCGAGTTGTTTCATCTGAA 921
QY      301 ThrProLysArgProLysValLysArgIleProGlnLysValGlnIleGluSerAsnThr 320
DB      922 AGCCAAAAACGGCAAAAGTGAACGAATTCCTCAAAAGTCAAAATCGAATCGAATACT 981
QY      321 ThrAlaSerThrThrGlnAsnSerGlyValArgValGluArgGluLeuAlaAlaGluArg 340
DB      982 ACAGCTTCAACGACACACAAACTCTGAGTGAAGTGAAGCGGAGCTAGCTGCTGAGAGA 1041
QY      341 ProThrValGlnAspAlaAsnPheHisArgHisGluGluGluArgGlyValLysLeu 360
DB      1042 CCAACTGTACAGACGCCCAATTTTTCACAGGCACGAAAGAGAGAGAGAGGTGTGAACCTT 1101
QY      361 GlyLeuGlyAspPheIlePheTyrSerValLeuLeuGlyLysAlaSerSerTyrPheAsp 380
DB      1102 GGTCTGGGCGACTTCATTTTCTACTCTGTTCTCTCTCGGCAAGGCTTCATCGTACTTTGAC 1161
QY      381 TrpAsnThrThrIleAlaCysTyrValAlaIleLeuIleGlyLeuCysPheThrLeuVal 400
DB      1162 TGGAAACACGACTATCGCTTGTATGTGGCCATTTCTTTCGGTCTCTGCTTCACTTCTGTC 1221
QY      401 LeuLeuAlaValPheLysArgAlaLeuProAlaLeuGlnPheProPheSerProAspSer 420
DB      1222 CTGCTCGCGCTCTTCAACGAGACTCCCGGCTCTGCAATTTTCCATTTTCTCCGAGCTCA 1281
QY      421 PhePheThrPheValProAlaGlySerSerProHisLeuLeuHisLysSerLeuLysSer 440
DB      1282 TTTTTTACTTTTGTACCGCTGATCATCATCCCCCATTTGTTTACACAGTCTCTCAAAAGT 1341
QY      441 ValTyrTyrIleAsnSerLeuPheLeuProPheLeuCysIleIleAsnPheSerIleIle 460
DB      1342 GTTATATATTAATTTCTCTGTTTTTGGCAATTTCTTGCATCATCACTTTTTCGATTATA 1401
QY      461 Ser 461
DB      1402 TCT 1404

```

RESULT 2

AAL47322 standard; DNA; 4137 BP.
 ID AAL47322; (first entry)
 AC AAL47322;
 DT 02-SEP-2002 (first entry)
 XX C elegans sel-12 gene promoter and regulatory regions.
 DE Sel-12; presenilin; neuronal disorder; familial Alzheimer's disease;
 KW amyloid precursor protein; APP; ds.
 XX Caenorhabditis elegans.
 OS US6376239-B1.
 PN 23-APR-2002.
 XX 04-APR-1997; 97US-0832867.
 XX 04-APR-1997; 97US-0832867.
 XX (ELEG-) ELEGNE GMBH.
 XX Baumeister R;
 XX WPI; 2002-478281/51.
 XX Isolated DNA molecule comprising promoter of the sel-12 gene from
 PT Caenorhabditis elegans operably linked to heterologous gene, directs
 PT expression in neural cells and is useful to develop drugs to treat
 PT neuronal disorders -
 XX Claim 1; Fig 3; 78pp; English.
 PS
 CC The present invention relates to DNA molecules comprising the promoter of
 CC the sel-12 gene from Caenorhabditis elegans operably linked to a
 CC heterologous DNA sequence encoding a protein of interest. The sequence
 CC can be used to develop drugs for the treatment, prevention or delay of a
 CC neuronal disorder. In particular, the neuronal disorder may be familial
 CC Alzheimer's disease. The present sequence is the C. elegans sel-12
 CC promoter.
 XX
 SQ Sequence 4137 BP; 1252 A; 770 C; 703 G; 1412 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.4e-196 Length: 4137
 Score: 2052.50 Matches: 459
 Percent Similarity: 58.82% Conservative: 1
 Best Local Similarity: 58.70% Mismatches: 1
 Query Match: 86.31% Indels: 322
 DB: 24 Gaps: 7
 US-09-043-944-1 (1-461) x AAL47322 (1-4137)
 QY 1 MetProSerThrArgArgGlnGlnGlnGlyGlyAlaAspAlaGluThrHis----- 18
 DB 960 ATGCTTCCACAGGACACACAGGAGGCGGAGGTGCAGATCGGAAACACATGTAAGT 1019
 QY 18 ----- 18
 DB 1020 TATTTAGACATTTTATTTTCTCAAGAACTAAATTTGTTAAATTTGCTACATGCTT 1079
 QY 19 -----ThrValTyGlyThrAsnLeuIleThrAsnArgAsnSerGlnGluAsnVa 37
 DB 1080 TCACACGGTTTACGGTACAAATCTGTATACAAATCGGAATAGCAAGACGAAATGT 1139
 QY 37 lValGluGluAlaGluLeuLysTyGlyAlaSerHisValIleHisLeuPheValProVa 57
 DB 1140 TGTGGAGAGCGGAGCTGAAATACGAGGATCTCAGTTTATTCATCTATTGTGCGGT 1199
 QY 57 lSerLeuCysMetAlaLeuValValPheThrMetAsnThrIleThrPheTySerGlnAs 77

Db 1200 GTCACTATGCATGGCTCTGGTTGTTTACGATGACACGATTACGTTTATATAGTCAAAA 1259
 QY 77 nAsnGlyArgHisLeu----- 82
 Db 1260 CAATGGAGGCAATTTGTAAAGTTTCTAAAGAAATTCATTGATTAATAATATTGATTTGTT 1319
 QY 83 -----LeuSer----- 84
 Db 1320 TTATCAATTTGCATCTGTCGCAATCGACTCTTTGTCTAGTCGCAAAATAATTTTGTGTCAGT 1379
 QY 84 ----- 84
 Db 1380 GCAAAATAATTTTGGTCAGTCGATTTGATTTATGGTCAGTCGATTTGCAAGTCGAGCT 1439
 QY 84 ----- 84
 Db 1440 TTAACATATTTTCGGTGGTTTAAATTTTACTCAATTTTCTATCAATATTTCTTTGGAAAAA 1499
 QY 84 ----- 84
 Db 1500 GTTGAAGATTTACTCTGGAAATTTTGGAAATAAATCTGTAATGGAATAATCAATCAACACA 1559
 QY 85 -----His-ProPheValArgGluThrAspSerIleValG 96
 Db 1560 AACTTTGAATATTTTTCAGACTATACACTCTCTTTTGTCCGGGAAACACACAGTAGTATCGTTG 1619
 QY 96 lLysGlyLeuMetSerLeuGlyAsnAlaLeuValMetLeuCysValValValLeuMetT 116
 Db 1620 AGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1679
 QY 116 hrValLeuLeuValPheTyGlySerPheTyGlySerPheTyGlySerPheTyGlySerPheTyGly 136
 Db 1680 CAGTTCTGCTGATTTTCTATAAATACAAAGTTTATTAAGCTTTATTCATGATGGCTTA 1739
 QY 136 lValSerSerPheLeuLeuPheLeuPheThrIleTyValGln----- 152
 Db 1740 TTGTGACGAGTTTCT 1799
 QY 153 -----GluValLeuLeuSerPheAspValS 161
 Db 1800 TATTACTATTCTCAATAAAATATCAATGTTGACAGAGAAGTTCTGAAAGTTTCTCGAATGATGT 1859
 QY 161 erProSerAlaLeuLeuValLeuPheGlyLeuGlyAsnTyGlyValLeuGlyMetMetC 181
 Db 1860 CTCCACGCGCACTATTTGGTTTGTGGACTGGGTAACTATGGAGTTCTCGGAATGATGT 1919
 QY 181 ysIleHisTrpLysGlyProLeuArgLeuGlnGlnPheTyLeuIleThrMetSerAla 201
 Db 1920 GTATACATTTGAAAGTCCATTTGGTCTGCAACAGTTCTACCTTATTACAAATGTTCTGCAC 1979
 QY 201 euMetAlaLeuValPheIleLysTyLeuProGluTrpThrValTrpPheValLeuPheV 221
 Db 1980 TAAATGGCTCTGCTCTTTATCAAGTACCTACCAGAATGGACTGTGTGGTTTGTGCTGTTG 2039
 QY 221 alIleSerValTrpAspLeuValAlaValLeuThrProLysGlyProLeuArgTyLeuV 241
 Db 2040 TTATCTCGGTTTGGATCTGTTTCCGGTCTCACACCAAAAGGACCAATTGAGATATTGG 2099
 QY 241 alGluThrAlaGlnGluArgAsnGluProIlePheProAlaLeuIleTySer----- 258
 Db 2100 TGGAACTGCACAGAGAGAACGAGCAATTTTCCGGCGCTGATTTTTCGTGTAAGT 2159
 QY 259 -----Ser-Gly 260
 Db 2160 TTCCTAATTTTGAATTAATTTATTCATGACGTTTCAAATTTCTAAACATTTTTCAGTCTGA 2219
 QY 261 ValIleTyProTyValLeuValThrAlaValGluAsnThrThrAspProAlaGluPro 280
 Db 2220 GTCATCTACCTACGTTCTTTGTTACTGAGTTGAAACACACAGACAGACCCCGGTGAACCG 2279
 QY 281 ThrSerSerAspSerAsn----- 286

Db 1697 GATGTTCTCTTT 1689

RESULT 8
AAT40028
ID AAT40028 standard; DNA; 2765 BP.
XX
XX AAT40028;
AC
XX
XX 25-MAR-2003 (updated)
DT 22-JUL-1997 (first entry)
DT
XX
XX Presenilin-1-1 wild type coding sequence.
DE
XX
XX Presenilin-1; human; hPS1-1; PS-2; integral membrane protein; AD;
KW familial Alzheimer's disease; cerebral haemorrhage; schizophrenia;
KW depression; antibody; gene expression modulator; therapy; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 249..1652
FT /*tag= a
FT /product= presenilin-1-1
XX
XX MO9634099-A2.
XX
XX 31-OCT-1996.
XX
XX 29-APR-1996; 96WO-CA00263.
XX
XX 31-JUL-1995; 95US-0509359.
XX 28-APR-1995; 95US-0431048.
XX 28-JUN-1995; 95US-0496841.
XX
XX (HSCR) HSC RES & DEV LP.
XX (UTOR) UNIV TORONTO GOVERNING COUNCIL.
XX
XX St Georgehslop PH, Fraser PE, Rommens JM;
XX WPI; 1996-497631/49.
XX P-PSDB; AAW05733.
XX
XX New presenilin genes - useful for diagnosis, therapy and drug
XX screening of familial Alzheimer's disease, cerebral disorders, etc.
XX
XX Claim 8; Page 128-130; 178pp; English.
XX
XX AAT40028 and AAT40029 represent the coding sequences for the two
XX different forms of wild type human presenilin-1 (PS-1). The form
XX represented by AAT40029 results from alternate splicing of the genomic
XX DNA sequence. AAT40031 represents the coding sequence for wild type human
XX PS-2. The presenilins are a family of highly conserved integral membrane
XX proteins with a common structural motif, common alternate splicing
XX patterns, and common mutational hot spot regions. Mutations in PS genes
XX are implicated in familial Alzheimer's disease (AD) and possibly other
XX diseases such as cerebral haemorrhage, schizophrenia, depression etc., so
XX detection of mutations in these sequences can be used for diagnosis of
XX these diseases. The encoded proteins, or vectors that express them or
XX containing antisense sequences, antibodies selective for mutant forms of
XX the encoded proteins (such as AAW05736) and modulators of PS gene
XX expression are potentially useful for treatment of AD etc. Transgenic
XX animals are useful as models for drug screening. The antibodies can also
XX be used e.g. for affinity purification and in immunoassays.
XX (Updated on 25-MAR-2003 to correct PI field.)
XX
XX SQ Sequence 2765 BP; 715 A; 624 C; 652 G; 773 T; 1 other;

Alignment Scores:
Pred. No.: 7,78e-90 Length: 2765
Score: 993.50 Matches: 216
Percent Similarity: 65.84% Conservative: 77
Best Local Similarity: 48.54% Mismatches: 104
Query Match: 41.78% Indels: 48

Db 1380 GGAGTAAACTTGATTTGGGAGATTTTCATTTCTACAGTGTCTGTTGGTAAAGCCTCA 1439
 Qy 377 SerTyrPhe-----AspTTPAsnThrThrIleAlaCysTyrValAlaIleLeuIleGly 394
 Db 1440 GCACACCGAGTGGAGACTGGACACACCATAGCCTGTTCGTAGCCATATTAATTTGGT 1499
 Qy 395 LeuCysPheThrLeuValLeuAlaValPheIysArgAlaLeuProAlaLeu-GlnPh 414
 Db 1500 TTGTGCTTACATTAATTTACTCTTGGCATTTTCAAGAAAGCATTTGCCAGCTCTTCCAAATC 1559
 Qy 414 eProPheSerProAspSerPhePheThrPheValProAlaGlySerSerProHisLeuLe 434
 Db 1560 TCCA--TCACCTTTGGGCTTGTTCCTTACTTTGCCACAGATATCTTTGACAGCCTTTT 1616
 Qy 434 uHisLysSerLeuLysSerValTyrTyrIleAsnSerLeuPhe-----LeuProPh 451
 Db 1617 ATGGACCATAGCATTCATCAATTTTATATCTAGCATATTTGCCGTTAGATCCCATG 1676
 Qy 451 eLeuCysIleIle 455
 Db 1677 GATGTTTCTTCTT 1689

RESULT 12

AAT59536
 ID AAT59536 standard; cDNA; 1750 BP.

XX AC AAT59536;

XX DT 07-MAY-1997 (first entry)

XX DE Human early onset Alzheimer's disease (EOAD) splice variant gene.

XX KW Early onset Alzheimer's disease; EOAD; neurodegenerative disease;

XX KW diagnosis; gene therapy; antisense; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 CDS 174..1565
 FT /*tag= a
 FT 248..249
 FT /*tag= b
 FT /label= VRSQ
 FT /note= "splice site"

XX FN W09703086-A1.

XX XX 30-JAN-1997.

XX XX 26-JUN-1996; 96WO-US11064.

XX XX 18-JUL-1995; 95US-0001501.

XX XX 13-JUL-1995; 95US-0001142.

XX XX (UYSF-) UNIV SOUTH FLORIDA.

XX XX Hardy JA;

XX XX WPI; 1997-118980/11.

XX XX P-PSDB; AAW11840.

XX XX Early onset Alzheimer's disease gene - useful for diagnosing a

XX XX pre-disposition to Alzheimer's disease

XX XX Claim 1; Fig 2; 44pp; English.

XX CC A full-length cDNA (AAT59536) of the early onset Alzheimer's disease
 CC (EOAD) splice variant gene sequence codes for a 463-amino acid
 CC polypeptide (AAW11839). Another full-length cDNA (AAT59535) of the
 CC EOAD gene codes for a 467-amino acid polypeptide (AAW11840). The 2
 CC sequences can be used to generate primers and probes for the
 CC diagnosis of predisposition to Alzheimer's disease, esp. EOAD.

CC They can also be used for prodn. of EOAD polypeptides in
 CC transformed host cells, and antisense sequences can be used for
 CC the treatment of EOAD.

XX SQ Sequence 1750 BP; 442 A; 389 C; 430 G; 478 T; 11 other;

Alignment Scores:

Score: 7.02e-90 Length: 1750
 Pred. No.: 991.00 Matches: 215
 Percent Similarity: 65.98% Conservative: 72
 Best Local Similarity: 49.43% Mismatches: 103
 Query Match: 41.67% Indels: 45
 DB: 18 Gaps: 8

US-09-043-944-1 (1-461) x AAT59536 (1-1750)

Qy 32 GlnGluAspGluAsnValValGluGluAlaGluLeuLysTyrGlyAlaSerHisValIle 51
 Db 351 GAGCAAGATGAGGAAGAAGATGAGAGCTGACATTTGAATATGGCCGACATGTGATC 410
 Qy 52 HisLeuPheValProValSerLeuCysMetAlaLeuValPheThrMetAsnThrIle 71
 Db 411 ATGCTCTTTGTCCTGTGACTCTCTGTCATGTGTGGTGGTGGCTACCATTAAGTCAGTC 470
 Qy 72 ThrPheTyrSerGlnAsnAsnGlyArgHisLeuLeuSerHisProPheValArgGluThr 91
 Db 471 AGCTTTTATACCGGAGGATGGGCAG---CTAATCTATATCCCATTCACAGAAATACC 527
 Qy 92 AspSerIleValGluLysGlyLeuMetSerLeuGlyAsnAlaLeuValMetLeuCysVal 111
 Db 528 GAGACTGTGGCCAGAGAGCCCTGCACCTCAATTCGAATGTGCCATCATGATCAGTGT 587
 Qy 112 ValValLeuMetThrValLeuLeuIleValPheTyrLysTyrLysPheTyrLysLeu 131
 Db 588 ATTGTGTGATGACTATCTCTGCTGTGCTTCTCTATAAATACAGGTGTCTATAAGTCATC 647
 Qy 132 HisGlyTyrLeuIleValSerSerPheLeuLeuPheLeuPheThrThrIleTyrVal 151
 Db 648 CATGCTGGCTTATATATCATCTCTATGTGTGCTGCTCTTTTTCATTCATTACTTG 707
 Qy 152 GlnGluValLeuLysSerPheAspValSerProSerAlaLeuLeuValLeuPheGlyLeu 171
 Db 708 GGGGAAGTGTTTAAACCTATAACGTTGCTGGGACTACATTACTGTGCACTCCTGATC 767
 Qy 172 GlyAsnTyrGlyValLeuGlyMetMetCysIleHisTyrLysGlyProLeuArgGluGln 191
 Db 768 TGGAAATTTGGTGTGGTGGATGATTTCCATCTCTCGAAGGTCCTTCGATCCAG 827
 Qy 192 GlnPheTyrLeuIleThrMetSerAlaLeuMetAlaLeuValPheIleLysTyrLeuPro 211
 Db 828 CAGGCATATCTCATTTATGATGATGCTCCCTCATGGCCCTGGTGTGTTATCAAGTACCTCCT 887
 Qy 212 GluTyrThrValTyrPheValLeuPheValIleSerValTyrAspLeuValAlaValLeu 231
 Db 888 GAATGGACTGCGTGCTCATCTTGGCTGTGATTTTCAGTATATGATTTAGTGGCTGTTTG 947
 Qy 232 ThrProLysGlyProLeuArgTyrLeuValGluThrAlaGlnGluArgAsnGluProIle 251
 Db 948 TGTCGGAAGGTCCACTTCGTATGCTGTGTGTAACAGCTCAGAGAGAAATGAACGCTT 1007
 Qy 252 PheProAlaLeuIleTyrSerSerGlyValIleTyrProTyrValLeuValThrAlaVal 271
 Db 1008 TTTCAGCTCTCATTTACTCTCTCAACATGGTGTGG-----TTGGTGAATATGCA 1058
 Qy 272 GluAsnThrThrAspProArgGluProThrSerSerAspSerAsnThrSerThrAlaPhe 291
 Db 1059 GAAGGAGACCCGAGAGCTCAAGAGGAGATATCCAAAATTC----- 1100
 Qy 292 ProGlyGluAlaSerCysSerSerGluThrProLysArgProLysValLysArgIlePro 311
 Db 1101 -----AAGTATAATGACAGAAAGCAGACAGAGG----- 1127
 Qy 312 GlnLysValGlnIleGluSerAsnThrThrAlaSerThrThrGlnAsnSerGlyValArg 331

Db 1071 GAAGGAGACCGGAGCTCAAGAGAGATGCCAAATATCC----- 1112
 Qy 292 ProGlyAlaSerCysSerSerGluThrProLysArgProLysValLysArgIlePro 311
 Db 1113 -----AAGTATAATGCAGAGAGACAGAAAGG----- 1139
 Qy 312 GlnLysValGlnIleGluSerAsnThrThrAlaSerThrThrGlnAsnSerGlyValArg 331
 Db 1140 -----GAGTCACAGACACTGTGTCAGAGATGATGATGGCGGTTCAGT 1184
 Qy 332 ValGlu-----ArgGluLeuAlaAlaGlu 339
 Db 1185 GAGGATGGGAAGCCAGAGGAGGAGCAGTCATCTAGGCGCTCATCGCTCTACACCTGAGTCA 1244
 Qy 340 ArgProThrValGlnAspAlaAsn-----PheHisArgHisGluGluGluArg 356
 Db 1245 CGAGCTCTGTCCAGGAACCTTCCAGCAGATCTCTGCTGCTGAGAGACCCAGAGGAAGG 1304
 Qy 357 GlyValLysLeuGlyLeuGlyAspPheIlePheThrSerValLeuLeuGlyLysAlaSer 376
 Db 1305 GGAGTAAACCTTGATGGGAGATTTCATTTCTACAGTGTCTGCTGTTGTTGTAAGCCTCA 1364
 Qy 377 SerTyrPhe-----AspTrpAsnThrThrIleAlaCysTyrValAlaIleLeuGly 394
 Db 1365 GCACAGCCAGTGGAGACTGGACACACCATAGCTGTTCGTAGCCATATTAATTTGGT 1424
 Qy 395 LeuCysPheThrLeuValLeuAlaValPheLysArgAlaLeuProAlaLeu-GlnPh 414
 Db 1425 TTGGCGCTTACATATTACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1484
 Qy 414 eProPheSerProAspSerPhePheThrPheValProAlaGlySerSerProHisLeu 434
 Db 1485 TCCA---TCACCTTTGGGCTGTTGTTCTTCTACTTTGCCACAGATTATCTGTACAGCCTTT 1541
 Qy 434 uHisLysSerLeuLysSerValTyrTyrIleAsnSerLeuPhe 448
 Db 1542 ATGGACCAATTAGCATTCATCAATTTTATATCTAGCATATTT 1584
 RESULT 14
 ID ABK83912
 XX ABK83912 standard; cDNA; 3056 BP.
 XX AC ABK83912;
 XX DT 14-AUG-2002 (first entry)
 XX DE Human cDNA differentially expressed in granulocytic cells #483.
 XX KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KW viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.
 XX OS Homo sapiens.
 XX FN WO200228999-A2.
 XX PD 11-APR-2002.
 XX PF 03-OCT-2001; 2001WO-US30821.
 XX PR 03-OCT-2000; 2000US-237189P.
 XX PA (GENE-) GENE LOGIC INC.
 XX PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 XX WPI; 2002-435328/46.
 XX

PT Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity -
 XX Claim 1; SEQ ID No 483; 114pp; English.
 XX The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) GA by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC or allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GA; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, ARDS, adult respiratory distress syndrome, renal
 CC reperfusion injury, asthma, thrombosis, cardiac reperfusion injury,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 3056 BP; 762 A; 688 C; 740 G; 866 T; 0 other;
 Alignment Scores:
 Pred. No.: 2,29e-89 Length: 3056
 Score: 989.50 Matches: 216
 Percent Similarity: 65.62% Conservatives: 76
 Best Local Similarity: 48.54% Mismatches: 105
 Query Match: 41.61% Indels: 48
 DB: 24 Gaps: 9
 US-09-043-944-1 (1-461) x ABK83912 (1-3056)
 QY 32 GlnGluAspGluAsnValValGluGluAlaGluLeuLysTyrGlyAlaSerHisValIle 51
 Db 731 GAGCAGATGAGGAGAGAGATGAGGAGCTGACATTTGAATATATGGCCCAAGCTGTGATC 790
 QY 52 HisLeuPheValProValSerLeuLysMetAlaLeuValValPheThrMetAsnThrIle 71
 Db 791 ATGCTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 850
 QY 72 ThrPheTyrSerGlnAsnAsnGlyArgHisLeuLeuSerHisProPheValArgGluThr 91
 Db 851 AGCTTTTATACCGGAG 907
 QY 92 AspSerIleValGluLysGlyLeuMetSerLeuGlyAlaAlaLeuValMetLeuCysVal 111
 Db 908 GAGACTGTGGCCGAG 967
 QY 112 ValValLeuMetThrValLeuLeuIleValPheTyrLysTyrLysPheTyrLysLeuIle 131

```

968 ATTGTTGTCATGATATCTCTCTGTTGTTCTGTATATAAATACAGTGTCTATAAGGTATC 1027
132 HisGlyTrpLeuIleValSerPheLeuLeuLeuPheLeuPheThrIleTyrVal 151
1028 CATGCTGGCTTATTATATCATCTATTATTGTTGTTCTTTTTCATTCAATTACTTG 1087
152 GlnGluValLeuLeuSerPheAspValSerProSerAlaLeuLeuValLeuPheGlyLeu 171
1088 GGGGAAGTGTTTAAACCTATATACGTTGCTGTGAGTACATTAATCTGTGCTCTGATC 1147
172 GlyAsnTyrGlyValLeuGlyMetCysIleHisTrpLysGlyProLeuArgLeuGln 191
1148 TGGAAATTTGGGTGGTGGGATGATTTCCATTCTCTGGAAGGTCCATCTCGACTCAG 1207
192 GlnPheTyrIleLeuThrMetSerAlaLeuMetAlaLeuValPheIleLysTyrLeuPro 211
1208 CAGGCATATCTCATATGATTAGTGCCTCATGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1267
212 GluTrpThrValTrpPheValLeuPheValIleSerValTrpAspLeuValAlaValLeu 231
1268 GAATGGACTGCTGGTGGCTCATCTCTGCTGTGATTTTCAGTATATGATTTAGTGGCTGTTTG 1327
232 ThrProLysGlyProLeuArgTyrLeuValGluThrAlaGlnGluArgAsnGluProIle 251
1328 TGTCGGAAGTCCACATTCGTATGCTGTTGAACAGCTCAGGAGAGAAATGAACGCTT 1387
252 PheProAlaLeuIleTyrSerSerGlyValIleTyrProTyrValLeuValThrAlaVal 271
1388 TTTCCAGCTCTCATTTACTCTCTCAACATGTTGTTG-----TTGGTGAATATGGCA 1438
272 GluAsnThrThrAspProArgGluProThrSerSerAspSerAsnThrSerThrAlaPhe 291
1439 GAAGGACCGCGAGAGCTCAAGAGAGATGCCAAAATTTCC----- 1480
292 ProGlyGluAlaSerCysSerSerGluThrProLysArgProLysValLysArgIlePro 311
1481 -----AAGTATAATGCAGAAAGCACAGAAAG----- 1507
312 GlnLysValGlnIleGluSerAsnThrThrAlaSerThrThrGlnAsnSerGlyValArg 331
1508 -----GAGTCACAGACACTGTTGCAGAGAAATGATGATGCGGGTTCAGT 1552
332 ValGlu-----ArgGluLeuAlaAlaGlu 339
1553 GAGGAATGGGAGCCAGAGGACAGTCATCTAGGGCTCATCTCTACACCTGAGTCA 1612
340 ArgProThrValGlnAspAlaAsn-----PheHisArgHisGluGluGluArg 356
1613 CGAGCTGCTGTCCAGGAACATTTCCAGCAGTATCTCTGCTGTAAGACCCAGAGAAAG 1672
357 GlyValLysLeuGlyLeuGlyAspPheIlePheTyrSerValLeuLeuGlyLysAlaSer 376
1673 GGAGTAAACTTGAATGGGAGATTTCAATTTCTACAGTGTCTGTTGTTGTAAGCCTCA 1732
377 SerTyrPhe-----AspTrpAsnThrThrIleAlaCysTyrValAlaIleLeuIleGly 394
1733 GCAACAGCCAGTGGAGACTGGAACACACCATAGCTGTTCTGTAGCAATTAATTTGGT 1792
395 LeuCysPheThrLeuValLeuAlaValPheLysArgAlaLeuProAlaLeuGlnPhe 414
1793 TTGTGCTCTTACATTTACTCTCTGCAATTTCAAGAAAGCATTCAGCTCTTCCAATC 1852
414 eProPheSerProAspSerPhePheThrPheValProAlaGlySerSerProHisLeuLe 434
1853 TCCA---TCACCTTTGGGCTGTTTCTTACTTTGTCACAGATTATCTTTGACAGCCTTTT 1909
434 uHisLysSerLeuLysSerValTyrTyrIleAsnSerLeuPhe-----LeuProPhe 451
1910 ATGGACCAATTAGCATTTCCATCAATTTTATATCTAGCATATTTGCGGTAGATCCCATG 1969
451 eLeuCysIleIle 455
      : : : : :

```

Db 1970 GATGTTTCTTCTT 1982

RESULT 15

AAT40029 standard; DNA; 3086 BP.

XX AAT40029;

XX 25-MAR-2003 (updated)

DT 23-JUL-1997 (first entry)

XX Presenilin-1-2, alternatively spliced coding sequence.

XX Presenilin-1; human; hPS1-1; hPS1-2; PS-2; integral membrane protein; AD; familial Alzheimer's disease; cerebral haemorrhage; schizophrenia; depression; antibody; gene expression modulator; therapy; ss.

OS Homo sapiens.

Key Location/Qualifiers
CDS 547..1948
FT /*tag= a
FT /product= presenilin-1-2

XX WO9634099-A2.

XX 31-OCT-1996.

XX 29-APR-1996; 96WO-CA00263.

XX 31-JUL-1995; 95US-0509359.

XX 28-APR-1995; 95US-0431048.

XX 28-JUN-1995; 95US-0496841.

XX (HSCR-) HSC RES & DEV LP.

XX (UTOR) UNIV TORONTO GOVERNING COUNCIL.

XX St Georgehyalop PH, Fraser PE, Rommens JM;

XX WPI; 1996-497631/49.

XX P-PSDB; AAW05734.

XX New presenilin genes - useful for diagnosis, therapy and drug screening of familial Alzheimer's disease, cerebral disorders, etc.

XX Claim 8; Page 132-134; 178pp; English.

XX AAT40028 and AAT40029 represent the coding sequences for the two different forms of wild type human presenilin-1 (PS-1). The form represented by this sequence results from alternate splicing of the genomic DNA. AAT40031 represents the coding sequence for wild type human PS-2. The presenilins are a family of highly conserved integral membrane proteins with a common structural motif, common alternate splicing patterns, and common mutational hot spot regions. Mutations in PS genes are implicated in familial Alzheimer's disease (AD) and possibly other diseases such as cerebral haemorrhage, schizophrenia, depression etc., so detection of mutations in these sequences can be used for diagnosis of these diseases. The encoded proteins, or vectors that express them or containing antisense sequences, antibodies selective for mutant forms of the encoded proteins (such as AAW05736) and modulators of PS gene expression are potentially useful for treatment of AD etc. Transgenic animals are useful as models for drug screening. The antibodies can also be used e.g. for affinity purification and in immunoassays.

XX (Updated on 25-MAR-2003 to correct PI field.)

SQ Sequence 3086 BP; 789 A; 688 C; 740 G; 866 T; 3 other;

Alignment Scores:

Pred. No.: 2,33e-89 Length: 3086
Score: 989.50 Matches: 216
Percent Similarity: 65.62% Conservative: 76
Best Local Similarity: 48.54% Mismatches: 105
Query Match: 41.61% Indels: 48

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 11, 2003, 07:41:57 ; Search time 2164 Seconds
(without alignments)

708.031 Million cell updates/sec

Title: US-09-043-944-1

Perfect score: 2378

Sequence: 1 MPSTRQQGGGADAETHV.....YYINSLFLFLCLINFIIS 461

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US09043944/runat.11122003.062402.3791/app.query.fasta.1.647
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -LISTS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LISTS=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORMEXT -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09043944 @CGN.1.107 @runat.11122003.062402.3791
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	994.5	41.8	2765	9	US-09-785-474-1	Sequence 1, Appli
2	994.5	41.8	2765	9	US-09-785-474-27	Sequence 27, Appli
3	993.5	41.8	2763	13	US-10-293-000-7	Sequence 7, Appli
4	993.5	41.8	2764	13	US-10-116-275-290	Sequence 230, App
5	993.5	41.8	2765	15	US-10-221-454-5	Sequence 5, Appli
6	989.5	41.6	2765	9	US-09-785-474-31	Sequence 31, Appli
7	984.5	41.4	2765	9	US-09-785-474-29	Sequence 29, Appli
8	979.5	41.2	2765	9	US-09-785-474-3	Sequence 3, Appli
9	979	41.2	1392	9	US-09-895-035-13	Sequence 13, Appli
10	979	41.2	1404	9	US-09-754-949-3	Sequence 3, Appli
11	979	41.2	1404	13	US-10-417-422-3	Sequence 25, Appli
12	967	40.7	2285	9	US-09-878-454A-25	Sequence 1, Appli
13	961	40.4	2236	13	US-10-180-781-1	Sequence 8, Appli
14	959	40.3	2236	13	US-10-293-000-8	Sequence 7, Appli
15	959	40.3	2236	15	US-10-221-254-7	Sequence 5, Appli
16	942.5	39.6	1346	9	US-09-754-949-5	Sequence 5, Appli
17	942.5	39.6	1346	13	US-10-417-422-5	Sequence 5, Appli
18	454	19.1	1362	10	US-09-938-842A-1045	Sequence 1045, Ap
19	337.5	14.2	473	13	US-10-180-781-14	Sequence 14, Appli
20	337.5	14.1	624	9	US-09-895-035-8	Sequence 8, Appli
21	272	11.4	493	9	US-09-895-035-3	Sequence 3, Appli
22	259.5	10.9	469	9	US-09-895-035-6	Sequence 6, Appli
23	259.5	10.9	816	9	US-09-895-035-2	Sequence 2, Appli
24	259	10.9	562	11	US-09-918-995-28131	Sequence 28131, A
25	248	10.4	411	13	US-10-029-386-4373	Sequence 4373, Ap
26	246	10.3	245	13	US-10-029-386-18073	Sequence 18073, A
27	236	9.9	2387	13	US-10-180-781-38	Sequence 38, Appli
28	202.5	8.5	356	9	US-09-895-035-5	Sequence 5, Appli
29	195.5	8.2	269	9	US-09-895-035-4	Sequence 4, Appli
30	186.5	7.8	332	9	US-09-925-299-454	Sequence 454, App
31	186.5	7.8	332	11	US-09-818-875-4212	Sequence 412, Ap
32	178	7.5	121	11	US-09-818-875-4213	Sequence 4213, Ap
33	178	7.5	121	11	US-09-818-875-4216	Sequence 4216, Ap
34	178	7.5	121	11	US-09-818-875-4217	Sequence 4217, Ap
35	178	7.5	121	11	US-09-818-875-4212	Sequence 4212, Ap
36	178	7.5	121	13	US-10-209-787-4213	Sequence 4213, Ap
37	178	7.5	121	13	US-10-209-787-4216	Sequence 4216, Ap
38	178	7.5	121	13	US-10-209-787-4217	Sequence 4217, Ap
39	178	7.5	121	13	US-10-209-787-4220	Sequence 4220, Ap
40	176	7.4	121	11	US-09-818-875-4221	Sequence 4221, Ap
41	176	7.4	121	11	US-10-209-787-4220	Sequence 4220, Ap
42	176	7.4	121	13	US-10-209-787-4221	Sequence 4221, Ap
43	176	7.4	121	13	US-10-209-787-4224	Sequence 4224, Ap
44	165	6.9	121	11	US-09-818-875-4225	Sequence 4225, Ap
45	165	6.9	121	11	US-09-818-875-4225	Sequence 4225, Ap

ALIGNMENTS

RESULT 1

US-09-785-474-1
; Sequence 1, Application US/09785474
; Patent No. US20010012626A1
; GENERAL INFORMATION:
; APPLICANT: TANZI, RUDOLPH
; WASCO, WILMA
; TITLE OF INVENTION: Genetic Alterations Related To Familial
; Alzheimer's Disease
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA: US/09/785, 474

US-09-878-454A-25

Alignment Scores:
Pred. No.: 2,42e-95 Length: 2285
Score: 967.00 Matches: 220
Percent Similarity: 61.95% Conservative: 60
Best Local Similarity: 48.67% Mismatches: 94
Query Match: 40.66% Indels: 78
DB: 9 Gaps: 12

US-09-043-944-1 (1-461) x US-09-878-454A-25 (1-2285)

QY	2	ProSerThrArgArgGlnGlnGluGly-----GlyGlyAlaAspAlaGluThrHisThr	19
Db	441	CCCAAGCGCGCCTCTCTGCAGAGGGGACGAGGCCCGCCAGAGGTAGGAG---AAATACT	497
QY	20	ValTyrGlyThrAsnLeuIleThrAsnArgAsnSerGlnGluAspGluAsn-----	36
Db	498	GCCAGTGGAGAAGCCAGGAGAACGAGAGGACGGTGAGGAGGACCCTGACCGCTATGTC	555
QY	37	-----ValValGluGluAlaGluLeuLysTyrGly	46
Db	558	TGTAGTGGGGTTCCCGGGCGCCGCCAGCGCTGGAGGAAGAGTGACCCCTCAAATAACGGA	617
QY	47	AlaSerHisValIleHisLeuPheValProValSerLeuCysMetAlaLeuValValPhe	66
Db	618	GCGAAGCATGTGATCATGCTGTTTGTGCTCTCATCTGTGCAATGATCGTGGTGTGATGCC	677
QY	67	ThrMetAsnThrIleThrPheTyrSerClnAsnAsnGlyArgHisLeuLeuSerHisPro	86
Db	678	ACCATCAAGTCTGTGGCTTCTCACAGAGAAGATGGACAG---CTCATCTACACGGCA	734
QY	87	PheValArgGluThrAspSerSileValcIuLysGlyLeuMetSerLeuGlyAsnAlaLeu	106
Db	735	TTCACTGAGGACACACCTCGTGGCGCAGCGCTCTCAACTCGTCTGTAACACCTC	794
QY	107	ValMetLeuCysValValLeuMetThrValLeuLeuIleValPheTyrLysTyrLys	126
Db	795	ATCATGATCAGCGTCATCGTGGTTATGACCATCTCTTGGTGGTGTCTACAAGTACCGC	854
QY	127	PheTyrLysLeuIleHisGlyTyrLeuIleValSerSerPheLeuLeuPheLeuPhe	146
Db	855	TGCTACAAGTTCACTCCAGGCTGGTTGATCATGTCCTCACTGATGTCGTCTGCTCTTC	914
QY	147	ThrThrIleTyrValGlnGluValLeuLysSerPheAspValSerProSerAlaLeuLeu	166
Db	915	ACCTATATCTACCTTGGGGAAGTGCTCAAGACCTACAATGTGGTCATGAGCTACCCCAC	974
QY	167	ValLeuPheGlyLeuGlyAsnTyrGlyValLeuGlyMetMetCysIleHisTyrLysGly	186
Db	975	CCTCTGTGCTACTGTCTGGAACCTTCGGGCGCAGTGGCATGGTGTGCATCCACTGGAAGGC	1034
QY	187	ProLeuArgLeuGlnGlnPheTyrLeuIleThrMetSerAlaLeuMetAlaLeuValPhe	206
Db	1035	CCCTGTGTGTCGACGAGGCGCTACCTCATCATGATCATGTAGCGCTCATGGCCCTAGTGTC	1094
QY	207	IleLysTyrLeuProGluThrThrValTyrPheValLeuPheValIleSerValTyrAsp	226
Db	1095	ATCAAGTACCTCCAGAGTGGTCCGGTGGGTCACTCTGGGGCCATCTCTGTGTATGAT	1154
QY	227	LeuValAlaValLeuThrProLysGlyProLeuArgTyrLeuValGluThrAlaGlnGlu	246
Db	1155	CTCGTGGCTGTGTGTCCCCAAGGCGCTCTCAGAAATGCTGGTAGAAACTGCCAGGAG	1214
QY	247	ArgAsnGluProIlePheProAlaLeuIleTyrSerSerGlyValIleTyrProTyrVal	266
Db	1215	AGAAATGAGCCCATATTCCCTGCCTGATATCATCTCCATGGTGTGGACGGTTGGC	1274
QY	267	LeuValThr-----AlaValcIuAsnThrThrAspProArg---	278
Db	1275	ATGGGAAGCTGGAACCCCTCTCTCAGGGTGGCCCTCCAGCTCCCTCACGACCGGAGATG	1334
QY	279	GluProThrSerSerAspSerAsnThrSerThrAlaPheProGluAlaSerCvsSer	298

Result No.	Score	Query Match	Length	DB	ID	Description
1	2052.5	86.3	4137	4	US-08-832-867-1	Sequence 1, Appl
2	994.5	41.8	2765	3	US-08-706-344C-1	Sequence 1, Appl
3	994.5	41.8	2765	3	US-08-706-344C-27	Sequence 27, Appl
4	993.5	41.8	1488	2	US-08-875-972-3	Sequence 3, Appl
5	993.5	41.8	2764	3	US-08-923-454A-9	Sequence 9, Appl
6	993.5	41.8	2764	4	US-08-832-867-2	Sequence 2, Appl
7	993.5	41.8	2765	3	US-08-888-077A-1	Sequence 1, Appl
8	993.5	41.8	2791	2	US-08-967-101-133	Sequence 133, App
9	993.5	41.8	2791	2	US-08-592-541-133	Sequence 133, App
10	993.5	41.8	2791	3	US-09-124-698-133	Sequence 133, App
11	993.5	41.8	2791	4	US-09-127-480-133	Sequence 133, App
12	993.5	41.8	2791	4	US-09-124-523-133	Sequence 133, App

18

18 *ov*

18 *ov*

Db 1020 TATTAGACATTTATTTTCTCAAGAACTAAATTTGTTAAATTTGCTACAAATGCTT 1079
 QY 19 -----ThrValTyrGlyThrAsnLeuLeuIleThrAsnArgAsnSerGlnGluAspGluAsnVa 37
 Db 1080 TCAGACCGCTTTACGGTACAAATCTGATACAAATCGGAATACCGAAGACGAAATGT 1139
 QY 37 lValGluGluAlaGluLeuLysTyrGlyValAsnHisValIleHisLeuPheValProVa 57
 Db 1140 TGTGAAGAGCGGAGCTGAAATACGAGCACTCTCACGTTATTCATCTATTTTGCCGGT 1199
 QY 57 lSerLeuCysMetAlaLeuValValPheThrMetAsnThrIleThrPheTyrSerGlnAs 77
 Db 1200 GTCACTATGTCATGCTGCTGGTGTGTTTATCATGATGAACACGATTACGTTTATAGTCAAAA 1259
 QY 77 nAsnGlyArgHisLeu-----LeuSer----- 82
 Db 1260 CAATGGAAGGCATTTGTAAGTTTCTAAAGATTCATTAAGATTCATTAAGATTTGATTTGTT 1319
 QY 83 ----- 84
 Db 1320 TTATCAATTTGTCATCTGTGCAATCGCACTCTTTGTGTCAGTGCATAATAATTTTGTGTCAGT 1379
 QY 84 ----- 84
 Db 1380 GCAAAATAATTTTGGTCAGTGCATTTGATTTATGTCAGTGCATTTGCAAGTCTGAGCT 1439
 QY 84 ----- 84
 Db 1440 TTAACTATTTTGGTGTGTTAAATTTTACTCAATTTTCTATCAATTTTCTTTTGGAAAAA 1499
 QY 84 ----- 84
 Db 1500 GTTGAAGATTTTACTCTGGAATTTTCGAAATAAACTGTAATTTGGAATAATCAATCAACACA 1559
 QY 85 -----His-PropheValArgGluThrAspSerIleValG 96
 Db 1560 AACTTTGAATTTTTCAGACTATACATCTCTTTTGTCTCGGAACACAGACAGTATCGTTG 1619
 QY 96 lLysGlyLeuMetSerLeuGlyAsnAlaLeuValMetLeuCysValValValLeuMetT 116
 Db 1620 AGAAGGATTTGATGTCACCTTGGAATGCTCTGTCATGTTGTCGTCGTTCTGATGA 1679
 QY 116 hrValLeuLeuIleValPheTyrLysTyrLysPheTyrLysLeuIleHisGlyTrpLeuI 136
 Db 1680 CAGTTCTGCTGATGTTTCTATAAATCAAGTTTATAAGCTTATTCATGATGAGTGGCTTA 1739
 QY 136 lValIserSerPheLeuLeuPhePheThrThrIleTyrValGln----- 152
 Db 1740 TTGTGAGCAGTTTCTCTCTCTTTCTTTTCTTATTCACATCAATCTATGTCAGTAAGTTGATA 1799
 QY 153 -----GluValLeuLysSerPheAspValS 161
 Db 1800 TATTACTATTCTCAFAAAATATCAATGTTGAGAGAAGTTCTGAAAAGTTTCGATGTGT 1859
 QY 161 exProSerAlaLeuLeuValLeuPheGlyLeuGlyAsnTyrGlyValLeuGlyMetMetC 181
 Db 1860 CTCCAGCGCACTATTGTTGTTTGTGACTGGTAACTATGAGTCTCTCGGAATGATGT 1919
 QY 181 yfIleHisTrpLysGlyProLeuArgLeuGlnGlnPheTyrLeuIleThrMetSerAlaL 201
 Db 1920 GTATACATTTGGAAGGTCATTCGCTGCTGCAACAGTTCTACCTATTATTAACAATGTCGCAC 1979
 QY 201 euMetAlaLeuValPheIleLysTyrLeuProGluTrpThrValTrpPheValLeuPheV 221
 Db 1980 TAAATGCTCTGCTCTTTATCAAGTACTACCAAGATGAGCTGTGTTGTTGCTGTTG 2039
 QY 221 alIleSerValTrpAspLeuValAlaValLeuThrProLysGlyProLeuArgTyrLeuV 241
 Db 2040 TTATCTCGGTTTGGATCTGTTGCTGCTCACACCAAAAGACCAATTGAGATATTGG 2099
 QY 241 aIGluThrAlaGlnGluArgAsnGluProIlePheProAlaLeuIleTyrSer----- 258

Db 2100 TGGAAACTGCACAGGAGAGAAACGAGCCAAATTTTCCGGCGCTGATTTTATTCGTGTAACT 2159
 QY 259 -----Ser-Gly 260
 Db 2160 TTCCTAATTTATGGAATTAATATTATCATGACGTTTCAAATTTCTAAACATTTTTCAGCTGGA 2219
 QY 261 ValIleTyrProTyrValLeuValThrAlaValGluAsnThrThrAspProArgGluPro 280
 Db 2220 GTCATCTATCCCTACGTTCTTGTACTGAGTTGAAACACGACAGACCCCGTGAACCG 2279
 QY 281 ThrSerSerAspSerAsn----- 286
 Db 2280 ACCTGCTCAGACTCAATATAGTGAATATCCACCTAAATTTTTCGAATTTTATTCGAAACT 2339
 QY 287 -----Thr-SerThrAlaPheProGlyGluAlaSerCysSerSerGluThrProLysAr 304
 Db 2340 AATTTTCAGCTTCTACAGCTTTTCTGGAGAGCGGAGTTGTTCTCATCTGAAACGCCAAACG 2399
 QY 304 gProLysValLysArgIleProGlnLysValGlnIleGluSerAsnThrThrAlaSerTh 324
 Db 2400 GCCAAAGTGAACGAATTCCTCAAAAGTGCATAATCGAATCGAATCTACAGCTTCAAC 2459
 QY 324 rThrGlnAsnSerGlyValArgValGluArgGluLeuAlaAlaGluArgProThrValG 344
 Db 2460 GACAAAACTCTGGAGTAAGGTGAGCGGAGCTAGCTCTGAGAGACCACTGTATCA 2519
 QY 344 nAspAlaAsnPheHisArgHisGluGluGluArg----- 356
 Db 2520 AGACGCCAATTTTACAGGACGACGAGAGAGAGAG- TGAGTGAAAAACGTCGTGAAAAA 2578
 QY 356 ----- 356
 Db 2579 GGGCAAAAGGGGATGATTTTTCGAAATTTTACTCGAACTTTCTCACTTCTTAACATAAT 2638
 QY 356 ----- 356
 Db 2639 GTTTTTTCTTGACAGCACAAACGAAATATTTGCCGCTCTACGTTCCGATCGAAATATCCC 2698
 QY 356 ----- 356
 Db 2699 CCTGCAATTTTCATTCGTTTTTTTTTCACTGTTTCAATTTTCTCAACTTTTGAAGAGCA 2758
 QY 356 ----- 356
 Db 2759 ATGCCGCCCACTCAGCTGAATATATTTGTTCAATTTAAAGTTTCAAAACTTTTTCAGTTAAT 2818
 QY 356 ----- 356
 Db 2819 AGATTCAAGAAAGATCTCAATAAACTTGCAGCTTGCACCTTGCCTAGTCACGAAAAA 2878
 QY 356 ----- 356
 Db 2879 AAGGATTTCTCAATAAACCTCTGTTTCATATTTTTTTTAAACAATAATTTTTCATCTCTTC 2938
 QY 357 -----GlyValLysLeuGlyLeuG 363
 Db 2939 ATATTTTGATATGTTTTCACAAAAAATGATTGACGAGGTTGTGAAACTTTGTTCTGGG 2998
 QY 363 YAspPheIlePheTyrSerValLeuLeuGlyLysAlaSerSerTyrPheAspTrpAsnTh 383
 Db 2999 CGACTTCATTTCTACTCTCTCTCTCGCAAGGCTTCATCGTACTTTGACTGGAACAC 3058
 QY 383 rThrIleAlaCysTyrValAlaIleLeuIleGlyLeuCysPheThrLeuValLeuLeuAl 403
 Db 3059 GACTATCGTCTGTTATGTGGCCATTTTATCGGCTCTCTGCTTCACTCTCTGCTGCTG 3118
 QY 403 aValPheLysArgAlaLeuProAlaLeu-GlnPheProPheSerProAspSerPhePheT 423
 Db 3119 CGTCTTCAACAGAGCACTCCCGCTCTGCCAATTTTCCATTTTCTCGGACTCATTTTTTA 3178
 QY 423 hrPheValProAlaGlySerSerProHisLeuLeuHisLysSerLeuLysSerValTyrT 443
 Db 3179 CTTTTGTACCCGCTGATCATCACCCCATTTGTTACACAAGTCTCTCAAAAGTGTATTAT 3238

Qy 443 yrIleAsnSerLeuPheLeuProPheLeuCysIleIleAsnPheSerIleIleSer 461
Db 3239 ATAATAATCTCTGTTTTTGGCATCTTTTGGCATCATCAACTTTTCGATTATATCT 3294

RESULT 2

US-08-706-344C-1
; Sequence 1, Application US/08706344C
; Patent No. 6248555
; GENERAL INFORMATION:
; APPLICANT: TANZI, RUDOLPH
; APPLICANT: MASCO, WILMA
; TITLE OF INVENTION: Genetic Alterations Related To Familial
; TITLE OF INVENTION: Alzheimer's Disease
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
;

Alignment Scores:			
Pred. No.:	2.31e-104	Length:	2765
Score:	994.50	Matches:	215
Percent Similarity:	65.17%	Conservative:	75
Best Local Similarity:	48.31%	Mismatches:	107
Query Match:	41.82%	Indels:	48
DB:	3	Gaps:	8
US-09-043-944-1 (1-461) x US-08-706-344C-1 (1-2765)			
Qy	32	GingluAspGluAsnValValtGluGluAlaGluLeuLysTyrglyAlaSerHisValIle	51
Db	438	GAGCAGATGAGGAGAGATGAGGAGCTGACATTGAATATGGCCGACATGTGATC	497
Qy	52	HisLeuPheValProValSerLeuCysMetAlaLeuValPheThrMetAsnThrIle	71
Db	498	ATGCCTTTTGCCCTGTGACTCTCTGATGGTGGTGGTCTGGTATCCATTAAAGTCAGTC	557

[illegible]

QY 434 uHisLysSerLeuLysSerValTyrTyrIleAsnSerLeuPhe-----LeuProPh 451
 Db 1617 ATGGACCAATTAGCATTCATCAATTTATATCTAGCATATTTCGGTTAGATCCCATG 1676
 QY 451 eLeuCysIleIle 455
 Db 1677 GATGTTCTCTT 1689

RESULT 3

US-08-706-344C-27
 ; Sequence 27, Application US/08706344C
 ; Patent No. 6248555
 ; GENERAL INFORMATION:
 ; APPLICANT: TANZI, RUDOLPH
 ; APPLICANT: WASCO, WILMA
 ; TITLE OF INVENTION: Genetic Alterations Related To Familial
 ; TITLE OF INVENTION: Alzheimer's Disease
 ; NUMBER OF SEQUENCES: 32
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
 ; STREET: 1100 NEW YORK AVENUE, SUITE 600
 ; CITY: WASHINGTON
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20005-3934

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/706,344C
 ; FILING DATE: 30-AUG-1996

; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/003,054
 ; FILING DATE: 31-AUG-1995

; ATTORNEY/AGENT INFORMATION:
 ; NAME: KIM, JUDITH U.

; REGISTRATION NUMBER: 40,679

; REFERENCE/DOCKET NUMBER: 0609.4180001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-371-2600

; TELEFAX: 202-371-2540

; INFORMATION FOR SEQ ID NO: 27:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2765 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 249..1649

US-08-706-344C-27

Alignment Scores:
 Pred. No.: 2,31e-104 Length: 2765
 Score: 994.50 Matches: 215
 Percent Similarity: 65.17% Conservative: 75
 Best Local Similarity: 48.31% Mismatches: 107
 Query Match: 41.82% Indels: 48
 DB: 3 Gaps: 8

US-09-043-944-1 (1-461) x US-08-706-344C-27 (1-2765)

QY 32 GlnGluAspGluAsnValValGluGluAlaGluLeuLysTyrGlyAlaSerHisValrle 51
 Db 438 GAGCAAGATGAGCAAGATGAGGAGCTGACATTGAAATATGGCCCAAGCATGTGATC 497

QY 52 HisLeuPheValProValSerLeuCysMetAlaLeuValPheThrMetAsnThrle 71
 Db 1500 TTGTGCTTACATTATTACTCTCTCCATTTTCAAGAAAGCATTCGACAGCTCTTCCAATC 1559

Db 498 ATGCTCTTTTGCCTCTGACTCTCTGCATGGTGGTGGTCTGCTGCTACCATTAAGTCAGTC 557
 QY 72 ThrPheTyrSerGlnAsnAsnGlyArgHisLeuLeuSerHisProPheValArgGluThr 91
 Db 558 AGCTTTTATACCGGAGGATGGGAG---CTAATCTATACCCCATTCACAGAAATACC 614
 QY 92 AspSerIleValGluLysGlyLeuMetSerLeuGlyAsnAlaLeuValMetLeuCysVal 111
 Db 615 GAGACTGTGGCCAGAGAGCCCTGCACCTCAATCTGATGCTGCCATCATGATCAGTGTC 674
 QY 112 ValValLeuMetThrValLeuLeuIleValPheTyrLysTyrLysPheTyrLysLeuIle 131
 Db 675 ATTGTTGTCATGACTACTCTCTCTGGTGTCTGTATAAATACAGGTGCTATAAGGTCATC 734
 QY 132 HisGlyTrpLeuIleValSerPheLeuLeuPheLeuPheThrThrIleTyrVal 151
 Db 735 CATGCTTGGCTTATATATCACTCTATTGTGTCTGTTCTTTTTCATCTCAITTACTTG 794
 QY 152 GlnGluValLeuLysSerPheAspValSerProSerAlaLeuLeuValLeuPheGlyLeu 171
 Db 795 GGGGAAGTGTATAAAACCTATAACGTTGCTGTGAGCTACATTACTTGTGCACCTCTGATC 854
 QY 172 GlyAsnTyrGlyValLeuGlyMetMetCysIleHisTyrLysGlyProLeuArgGln 191
 Db 855 TGGAAATTTGGTGTGGGAGATGATTCCATTCACTGGAAGGTCCACTTGCAGCTCCAG 914
 QY 192 GlnPheTyrLeuIleThrMetSerAlaLeuMetAlaLeuValPheIleLysTyrLeuPro 211
 Db 915 CAGGCATATCTCATTTATGATGATGAGTGCCTCATGCCCTGGTGTATCAAGTACCTCCCT 974
 QY 212 GluTrpThrValTrpPheValLeuPheValIleSerValTrpAspLeuValAlaValLeu 231
 Db 975 GAATGAGACTCGGTGGCTCATCTTGGCTGTGATTTCAATATATGATTTAGTGGCTGTTG 1034
 QY 232 ThrProLysGlyProLeuArgTyrLeuValGluThrAlaGlnGluArgAsnGluProIle 251
 Db 1035 CGTCCGAAAGTCCACTTCGTATGCTGGTGTGAACAGCTCAGAGAGAATGAAAGCCTT 1094
 QY 252 PheProAlaLeuIleTyrSerGlyValIleTyrProTyrValLeuValThrAlaVal 271
 Db 1095 TTTCCAGCTCTCATTTTACTCTCTCAACAATGTTGTGG-----TTGGTGAATATGGCA 1145
 QY 272 GluAsnThrThrAspProArgGluProThrSerSerAspSerAsnThrSerThrAlaPhe 291
 Db 1146 GAAGAGACCCGGAAGCTCAAGGAGAGTATCCAAAATTTCC----- 1187
 QY 292 ProGlyGluAlaSerCysSerSerGluThrProLysArgProLysValLysArgIlePro 311
 Db 1188 -----AAGCATATATGCA 1199
 QY 312 GlnLysValGlnIleGluSerAsnThrThrAlaSerThrThrGlnAsnSerGlyValArg 331
 Db 1200 GAAAGCACAGAAAGGAGTACAAAGACACTGTTTGACAGAAATGATGATGGCGGTTCAAT 1259
 QY 332 ValGlu-----ArgGluLeuAlaAlaGlu 339
 Db 1260 GAGGAATGGGAAGCCAGAGGAGACACTCATCTAGGCGCTCATCGCTCTACACTGAGTCA 1319
 QY 340 ArgProThrValGlnAspAlaAsn-----PheHisArgHisGluGluGluArg 356
 Db 1320 CGAGCTGCTGTCCAGGAACCTTTCCAGCAGTATCTCGCTGGTGAAGACCCAGAGAAAGG 1379
 QY 357 GlyValLysLeuGlyLeuGlyAspPheIlePheTyrSerValLeuLeuGlyLysAlaSer 376
 Db 1380 GGAGTAAACTGGATTGGAGATTTCATTTTCTACAGTGTCTGTTGGTAAAGCTCA 1439
 QY 377 SerTyrPhe-----AspTrpAsnThrThrIleAlaCysTyrValAlaIleLeuIleGly 394
 Db 1440 GCAACAGCCAGTGGAGACTGGAACAACAACATAGCTGTTTCGTAGCCATATTAATTTGTT 1499
 QY 395 LeuCysPheThrLeuValLeuLeuAlaValPheLysArgAlaLeuProAlaLeuGlnPh 414
 Db 1500 TTGTGCTTACATTATTACTCTCTCCATTTTCAAGAAAGCATTCGACAGCTCTTCCAATC 1559

QY 434 uHisLysSerLeuLysSerValTyrTrpIleAsnSerLeuPhe-----LeuProPh 451
 Db 1617 ATGACCAATTAGCATTCATCAATTTATATCTAGCATATTTCCGGTTAGATCCCATG 1676
 QY 451 eLeuCysIlelle 455
 Db 1677 GATGTTTCTCTT 1689
 RESULT 14
 US-08-496-841C-133
 ; Sequence 133, Application US/08496841C
 ; Patent No. 6210919
 ; GENERAL INFORMATION:
 ; APPLICANT: ST. GEORGE-HYSLOP, PETER H
 ; ROMMENS, JOHANNA M
 ; FRASER, PAUL E
 ; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
 ; NUMBER OF SEQUENCES: 175
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby, PC
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/496,841C
 ; FILING DATE: 28-Jun-1995
 ; CLASSIFICATION: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Paul F. Pehlner, Ph.D.
 ; REGISTRATION NUMBER: 35,135
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 527-7700
 ; TELEFAX: (212) 753-6237
 ; INFORMATION FOR SEQ ID NO: 133:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2792 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 133:
 US-08-496-841C-133
 Alignment Scores:
 Pred. No.: 3,06e-104 Length: 2792
 Score: 993.50 Matches: 216
 Percent Similarity: 65.84% Conservative: 77
 Best Local Similarity: 48.54% Mismatches: 104
 Query Match: 41.78% Indels: 48
 Gaps: 9
 US-09-043-944-1 (1-461) x US-08-496-841C-133 (1-2792)
 QY 32 GlnGluAspGluAsnValValGluGluAlaGluLeuLysTyrGlyAlaSerHisValile 51
 Db 438 GAGCAAGATGAGGAAGAAGATGAGGAGCTGACATTGAAATATGCGCGCAAGCATGTGATC 497
 QY 52 HisLeuPheValProValSerLeuCysMetAlaLeuValPheThrMetAsnThrIle 71
 Db 498 ATGCTCTTTGTCCTGTGACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 557
 QY 72 ThrPheTyrSerGlnAsnAspGlyArgHisLeuLeuSerHisProPheValArgGluThr 91
 Db 558 AGCTTTTATACCCGGAAGATGGGAG---CTAATCTATACCCCATTTACAGAGAATACC 614

QY 92 AspSerIleValGluLysGlyLeuMetSerLeuGlyAsnAlaLeuValMetLeuCysVal 111
 Db 615 GAGACTGTGGCCAGAGAGCCCTGCACTCAATTCCTGAATGCTGCCATCATGATCAGTGC 674
 QY 112 ValValLeuMetThrValLeuLeuIleValPheTyrLysTyrLysPheTyrLysLeu 131
 Db 675 ATTGTGTGTCATGACTATCTCTGCTGGTGTCTGTATAAATACAGGTGTATAGGTGCATC 734
 QY 132 HisGlyTyrLeuIleValSerSerPheLeuLeuLeuPheLeuPheThrThrIleTyrVal 151
 Db 735 CATGCTGGCTTATTATATCATCTCTATGTGTGCTGCTCTTTTTCATTCTATTACTTG 794
 QY 152 GlnGluValLeuLysSerPheAspValSerProSerAlaLeuValLeuValPheGlyLeu 171
 Db 795 GGGGAAGTGTTTAAACCTATAACCTGTCTGGGACTACATTACTGCTGTCACCTCCTGATC 854
 QY 172 GlyAsnTyrGlyValLeuGlyMetMetCysIleHisTyrLysGlyProLeuArgGln 191
 Db 855 TGGAAATTTTGGTGTGGTGGGAATGATTTCCATCTACTGGAAAGGTCCACTTCGACTCCAG 914
 QY 192 GlnPheTyrLeuIleThrMetSerAlaLeuMetAlaLeuValPheIleLysTyrLeuPro 211
 Db 915 CAGGCATATCTCATTTATGATTAGTCCCTCATGGCCCTGGTGTATTATCAAGTACCTCCCT 974
 QY 212 GluTyrThrValTyrPheValLeuPheValIleSerValTyrAspLeuValAlaValLeu 231
 Db 975 GAATGACTGCTGCTGCTCATCTTGGCTGTGATTTTCAGTATATGATTAGTGTGCTGTTG 1034
 QY 232 ThrProLysGlyProLeuArgTyrLeuValGluThrAlaGlnGluArgAsnGluProIle 251
 Db 1035 TGTCCGAAGGTCCACTTCTGCTGCTGTTGTTGAAACAGCTCAGAGAGAAATGAAAGCTT 1094
 QY 252 PheProAlaLeuIleTyrSerSerGlyValIleTyrProTyrValLeuThrAlaVal 271
 Db 1095 TTTCCAGCTCTCATTTACTCTCTCAACAAATGGGTGG-----TTGTGAATATGGCA 1145
 QY 272 GluAsnThrThrAspProArgGluProThrSerSerAspSerAsnThrSerThrAlaPhe 291
 Db 1146 GAAGGAGACCCGGAAGCTCAAGGAGAGATATCCAAAAATTC----- 1187
 QY 292 ProGlyGluAlaSerCysSerSerGluThrProLysArgProLysValLysArgIlePro 311
 Db 1188 -----AAGTATAATGCAGAAAGCAGAAAGG----- 1214
 QY 312 GlnLysValGlnIleGluSerAsnThrThrAlaSerThrThrGluAsnSerGlyValArg 331
 Db 1215 -----GAGTCACAGACACTGTTGCAGAGAATGATGATGGGGGTTCAGT 1259
 QY 332 ValGlu----- 339
 Db 1260 GAGGATGGGAAGCCCGAGGAGCAGTCACTAGGGCCCTCATCGCTCTACCTCAGTCA 1319
 QY 340 ArgProThrValGlnAspAlaAsn-----PheHisArgHisGluGluGluArg 356
 Db 1320 CGAGCTGTGCTCCAGGAACCTTTCCAGCAGATATCTCGCTGGTGAAGACCCAGAGAAAGG 1379
 QY 357 GlyValLysLeuGlyLysPheIlePheTyrSerValLeuLeuGlyLysAlaSer 376
 Db 1380 GGAGTAAACCTTGGATTGGGAGATTTCATTTCTACAGTGTCTGTTGGTGAAGACCTCA 1439
 QY 377 SerTyrPhe-----AspTyrAsnThrThrIleAlaCysTyrValAlaIleLeuIleGly 394
 Db 1440 GCAACAGCCAGTGGAGACTGGGAACAACCATAGCTCTTTTCGTAGCCATATTAATGTT 1499
 QY 395 LeuCysPheThrLeuValLeuLeuAlaValPheLysArgAlaLeuProAlaLeu-GlnPh 414
 Db 1500 TTGTGCTTTACATTATTACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1559
 QY 414 eProPheSerProAspSerPhePheThrPheValProAlaGlySerSerProHisLeuLe 434
 Db 1560 TCCA---TCACCTTTGGCTGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1616
 QY 434 uHisLysSerLeuLysSerValTyrTyrIleAsnSerLeuPhe-----LeuProPh 451

```
Db 1617 ATGACCAATAGCATTCATCAATTTATATCTAGCATATTTCGGTTAGAAATCCCATG 1676
QY 451 eLeuCysIleile 455
Db 1677 GATGTTCTTCTT 1689

RESULT 15
US-08-670-964-3
; Sequence 3, Application US/08670964
; Patent No. 6010874
; GENERAL INFORMATION:
; APPLICANT: Hardy, John A.
; TITLE OF INVENTION: EARLY ONSET ALZHEIMER'S DISEASE
; TITLE OF INVENTION: GENE AND GENE PRODUCTS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road - UW2220; P.O. Box 15
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/670,964
; FILING DATE: 26-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,142
; FILING DATE: 13-JUL-1995
; APPLICATION NUMBER: 60/001,501
; FILING DATE: 18-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: P50358
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1750 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-670-964-3

Alignment Scores:
Pred. No.: 2,68e-104 Length: 1750
Score: 991.00 Matches: 215
Percent Similarity: 65.98% Conservative: 72
Best Local Similarity: 49.43% Mismatches: 103
Query Match: 41.67% Indels: 45
DB: 3 Gaps: 8

US-09-043-944-1 (1-461) x US-08-670-964-3 (1-1750)
QY 32 GlnGluAspGluAsnValValGluGluAlaGluLeuLysTyrGlyAlaSerHisValile 51
Db 351 GAGCAAGATGAGGAGAGATGAGGAGCTGACATTGAAATATGGCCCAAGCATGTGATC 410
QY 52 HisLeuPheValProValSerLeuCysMetAlaLeuValPheThrMetAsnThrIle 71
Db 411 ATGCTCTTTGCTGACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 470
QY 72 ThrPheTyrSerGlnAsnAsnGlyArgHisLeuLeuSerHisProPheValArgGluThr 91
```

```
Db 471 AGCTTTTATACCCGGAAGGATGGGCAG---CTAATCTATACCCCAATTCACAGAGAATACC 527
QY 92 AspSerIleValGluLysGlyLeuMetSerLeuGlyAsnAlaLeuValMetLeuCysVal 111
Db 528 GAGACTGTGGCCAGAGAGCCCTGCCTCAATCTGAATGCTGCCATCATGATCAGTGTGTC 587
QY 112 ValValLeuMetThrValLeuLeuIleValPheTyrLysTyrLysPheTyrLysLeuLeu 131
Db 588 ATTTGTTGTCATGACTATCTCTCTCTGTTCTGTATAAATACAGGTCTATAAGGTCTATC 647
QY 132 HisGlyTyrLeuLeuValSerPheLeuLeuLeuPheLeuPheThrThrIleTyrVal 151
Db 648 CATGCTGCTGCTTATATATCATCTCTATTGTTCTGTTCTTTTTCATTCATTTACTTG 707
QY 152 GlnGluValLeuLysSerPheAspValSerProSerAlaLeuLeuValLeuPheGlyLeu 171
Db 708 GGGGAAGTGTTTAAACCTATAAGTGTGCTGTGAGTACATTTACTGTTGACTCTCTGATC 767
QY 172 GlyAsnTyrGlyValLeuGlyMetMetCysIleHisTyrLysGlyProLeuArgLeuGln 191
Db 768 TGGAAATTTTGGTGTGGTGGGATGATTTCCATTCACCTGGAAAGGTCCACTTCGACTCCAG 827
QY 192 GlnPheTyrLeuLeuThrMetSerAlaLeuMetAlaLeuValPheLeuLysTyrLeuPro 211
Db 828 CAGGCATATCTCATTTATGATTAGTGCCTCATGGCCCTGCTGTTTATCAAGTACCTCCCT 887
QY 212 GluTyrThrValTyrPheValLeuPheValIleSerValTyrPheValAlaValLeu 231
Db 888 GAATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 947
QY 232 ThrProLysGlyProLeuArgTyrLeuValGluThrAlaGlnGluArgAsnGluProIle 251
Db 948 TGTCCGAAAGTCCACTTCGTATGCTGTTGTTGAAACAGCTCAGGAGAGAAATGAACGCTT 1007
QY 252 PheProAlaLeuLeuTyrSerSerGlyValIleTyrProTyrValLeuValThrAlaVal 271
Db 1008 TTTCCAGCTCTCATTTACTCTCTCAACAATGGTGTGG-----TTGGTGAATATGGCA 1058
QY 272 GluAsnThrThrAspProArgGluProThrSerSerAspSerAsnThrSerThrAlaPhe 291
Db 1059 GAAGGAGAGCCCGAAGCTCAAGGAGAGATATCAAAAATTC----- 1100
QY 292 ProGlyGluAlaSerCysSerSerGluThrProLysArgProLysValLysArgIlePro 311
Db 1101 -----AAGTATAATGCAAGAAAGCAGCAAGAAAG----- 1127
QY 312 GlnLysValGlnIleGluSerAsnThrThrAlaSerThrThrGlnAsnSerGlyValArg 331
Db 1128 -----GAGTCACAAGACACTGTTTGCAGAGAATGATGCGCGGTTCAGT 1172
QY 332 ValGlu-----ArgGluLeuAlaAlaGlu 339
Db 1173 GAGGAATGGGAAGCCAGAGGAGACAGTCATCTAGGCGCTCATCGCTCTACACCTGAGTCA 1232
QY 340 ArgProThrValGlnAspAlaAsn-----PheHisArgHisGluGluGluArg 356
Db 1233 CGAGCTGCTGTCCAGGAACCTTCCAGCAGTATCTCGCTGGTGAAGACCCAGAGGAAAGG 1292
QY 357 GlyValLysLeuGlyLeuGlyAspPheIlePheTyrSerValLeuLeuGlyLysAlaSer 376
Db 1293 GGAGTAAACTTGGATTGGGAGATTTCATTTTCTACAGTGTTCGTGGTAAAGGCTCA 1352
QY 377 SerTyrPhe-----AspTyrAsnThrThrIleAlaCysTyrValAlaIleLeuIleGly 394
Db 1353 GCAACAGCAGCTGGAGACTGGACACCAACCATAGCTGTTTGTAGCCCATATTAATTTGT 1412
QY 395 LeuCysPheThrLeuValLeuLeuAlaValPheLysArgAlaLeuProAlaLeu-GlnPhe 414
Db 1413 TTGTGCTTACATTTACTCTTGCATTTTCAAGAAAGCATTTGCCAGCTCTTCCATC 1472
QY 414 eProPheSerProAspSerPhePheThrPheValProAlaGlySerSerProHisLeuLeu 434
```


Db 1473 TCCA---TCACCTTTGGGCTTGTTTTCTACTTTGCCACAGATTATCTTGTACAGCCTTTT 1529
 Qy 434 uHisLysSerLeuLysSerValTyrTyrIleAsnSerLeuPhe 448
 Db 1530 ATGGACCAATTAGCATTCCTCAATTTTATATCTTAGCATATTT 1572

Search completed: December 11, 2003, 07:43:07
 Job time : 122 secs